

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: September 17, 2005, 08:52:14 ; Search time 178 Seconds
(without alignments)
1556.376 Million cell updates/sec

Title: US-10-615-659-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCSSRR.....LRGLKTAEGALRPPPGCKGS 541

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1753.5	61.4	781	2 Q8CON7	Q8CON7 mus musculus
2	1726.5	60.5	518	2 Q8C0V2	Q8C0V2 mus musculus
3	1052	36.9	704	2 Q8BV51	Q8BV51 mus musculus
4	1043.5	36.6	744	2 Q9H876	Q9H876 homo sapien
5	981.5	34.4	434	2 Q8NDN8	Q8NDN8 homo sapien
6	957.5	33.5	352	2 Q6AWA3	Q6AWA3 homo sapien
7	956.5	33.5	352	1 TTL3 HUMAN	Q9Y477 homo sapien
8	834	29.2	266	2 Q922T0	Q922T0 mus musculus
9	746.5	26.2	580	2 Q7Q156	Q7Q156 anopheles g
10	722.5	25.3	992	2 Q9VM91	Q9VM91 drosophila
11	536.5	18.8	719	2 Q9VM92	Q9VM92 drosophila
12	524.5	18.4	281	2 Q6ZU95	Q6ZU95 homo sapien
13	385	13.5	101	2 Q96G98	Q96G98 homo sapien
14	345.5	12.1	461	2 Q641W7	Q641W7 rattus norv
15	345	12.1	464	2 Q9D570	Q9D570 mus musculus
16	335	11.7	1075	2 Q7QT13	Q7QT13 giardia lam
17	331.5	11.6	794	2 Q6BFH6	Q6BFH6 paramecium
18	328	11.5	496	2 Q961I9	Q961I9 drosophila
19	328	11.5	496	2 Q9VX74	Q9VX74 drosophila
20	314	11.0	403	2 Q7Q2C8	Q7Q2C8 giardia lam
21	311	10.9	673	2 Q6ZV70	Q6ZV70 homo sapien
22	307	10.8	559	2 Q8CFV5	Q8CFV5 mus musculus
23	307	10.8	1333	2 Q8CHB8	Q8CHB8 mus musculus
24	303.5	10.6	1048	2 Q9PLV5	Q9PLV5 homo sapien
25	303.5	10.6	1226	2 Q9UPZ4	Q9UPZ4 homo sapien
26	303.5	10.6	1277	2 Q6EMB2	Q6EMB2 homo sapien
27	301	10.5	438	2 Q8C125	Q8C125 mus musculus
28	299.5	10.5	423	1 TTL1 HUMAN	Q95922 homo sapien
29	298.5	10.5	423	1 TTL1 MOUSE	Q91V51 mus musculus
30	297.5	10.4	827	2 Q9VKI9	Q9VKI9 drosophila
31	297.5	10.4	828	2 Q8IGW4	Q8IGW4 drosophila

32	297.5	10.4	1070	2 Q8IPB2	Q8IPB2 drosophila
33	297.5	10.4	1295	2 Q6EEF3	Q6EEF3 cercopithec
34	292.5	10.2	440	2 Q7RKT2	Q7RKT2 plasmodium
35	291	10.2	541	2 Q8T417	Q8T417 drosophila
36	291	10.2	982	2 Q960F9	Q960F9 drosophila
37	291	10.2	989	2 Q9QV6	Q9QV6 drosophila
38	285.5	10.0	310	2 Q8I3W7	Q8I3W7 plasmodium
39	285	10.0	661	2 Q7QVA6	Q7QVA6 giardia lam
40	282.5	9.9	487	2 Q9V297	Q9V297 drosophila
41	280.5	9.8	827	2 Q7PVY2	Q7PVY2 anopheles g
42	271.5	9.5	592	1 TTL2 HUMAN	Q9BW77 homo sapien
43	270	9.5	917	2 Q9VXK5	Q9VXK5 drosophila
44	269	9.4	1060	2 Q7QAD8	Q7QAD8 anopheles g
45	265.5	9.3	436	2 Q7QZF0	Q7QZF0 giardia lam

ALIGNMENTS

RESULT 1

ID	Q8CON7	PRELIMINARY;	PRT;	781 AA.
AC	Q8CON7;			
DT	01-MAR-2003 (T-EMBLrel. 23, Created)			
DT	01-MAR-2003 (T-EMBLrel. 23, Last sequence update)			
DT	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)			
DE	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933401B17 product:hypothetical Tubulin-tyrosine ligase containing protein, full insert sequence.			
GN	Name=1700019P01Rik;			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;			
RX	Carninci P., Hayashizaki Y.;			
RT	"High-efficiency full-length cDNA cloning.";			
RL	Meth. Enzymol. 303:19-44(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	MEDLINE=21085660; PubMed=1217851; DOI=10.1038/35055500;			
RA	RIKEN FANTOM Consortium;			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	The FANTOM Consortium;			
RA	The RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";			
RL	Nature 420:563-573 (2002).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;			
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;			
RT	"Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";			
RL	Genome Res. 10:1617-1630(2000).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Testis;			
RC	MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;			
RA	Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto S., Ikegami T., Kashiwagi K.,			

RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishii K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tegami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK029745; BAC26595.1; -;
DR MGD; MGI:1922902; 1700019P01Rik.
DR GO: GO:0016874; F:ligase activity; IEA.
DR GO: GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO: GO:0006464; P:protein modification; IEA.
DR InterPro: IPR004344; Tub_tyr_lygase.
DR Pfam: PF03133; TTL; 1.
KW Hypothetical protein; Ligase.
SQ SEQUENCE 518 AA; 59318 MW; 540C045659FAC0D6 CRC64;

Query Match 36.5%; Score 1726.5; DB 2; Length 518;
Best Local Similarity 71.4%; Pred. No. 1.1e-110;
Matches 325; Conservative 57; Mismatches 64; Indels 9; Gaps 4;

Qy 1 MASSILKVVVSHQS-CRSRSRSPKRDQREAGSSDLSSRODAENAEAKLGLPGQLVDIA 59
Db VAASILKVVVHLQNYC---SKVKGSKKEBAKNSDPSPKDPENPDILKPLSLSGQVVDTA 115
Qy 60 CKVCOAYLGOLEHEDIDTSADAVEDLTEAWEDLTQOYYSILVHGDAPISNRYFSQCA 119
Db 116 CKVCOAYLGOLEHEDIDVSEASTEALSEEBWNDLTQOYYSILVHGNASITDSKSYFAQCA 175
Qy 120 LNRITSVNPTQTDIGLRNTWIIPKAASGRDVCMDRVEEILEAAADHPLSRUNKWV 179
Db 176 LLSKISSVNQTELDIGLRNTWIIPKAASGRDVCMDRVENTLSLVAADSQYTKONKWV 235
Qy 180 VQYIETPLICDTKFDIRQWFLVTDNPLTIWPKESYLFSTQPSLSKLDSAIHLN 239
Db 236 VQYIETPLIYDTKFDIRQWFLVTDNPLTIWPKESYLFSTQPSLSKLDSAIHLN 295
Qy 240 NAVQYLNKDVGRSPLPAHNMVTSFOEVLQCRGAVGWSVIVPSMKKATAHMKVA 299
Db 296 NSTQRLKNDKERSPLLPCHNMVTSFOEVLQCRGAVGWSVIVPSMKKRAVTNMRVA 355
Qy 300 QDVHVEPKNSFELYGADFLVGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVEDTIKVV 359
Db 356 QDVHVEAKNSFELYGADFLVGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVEDTIKVV 415
Qy 360 ----DRSCDIGNFELLWRQVPEPPPPSGSDLCVAGSVRRARRQVLPVNCMLKASASLLD 415
Db 416 DRKLDRCNDIGNFELLWRQVPEPPPPSGSDLCVAGSVRRARRQVLPVNCMLKASASLLD 475
Qy 416 AQLPKARGPSAMPDPAGPPSPALQRLGLKEEKG 450
Db 476 APP-KVRSARALMETVIRPRTTTRQDWKREEAKG 509

RESULT 3
Q8BV51 ID Q8BV51 PRELIMINARY; PRT; 704 AA.
AC Q8BV51;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus 3 days neonate thymus CDNA, RIKEN full-length enriched
DE library, clone:A630053H17 product:HTTL PROTEIN homolog.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.,
RT "High-efficiency full-length cDNA cloning.";

Meth. Enzymol. 303:19-44 (1999).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630 (2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771 (2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka I., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tegami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK080321; BAC37878.1; -;
DR GO: GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO: GO:0006464; P:protein modification; IEA.
DR InterPro: IPR004344; Tub_tyr_lygase.
DR Pfam: PF03133; TTL; 1.
SQ SEQUENCE 704 AA; 79080 MW; 3FAD889C1DB5CF7D CRC64;

Query Match 36.9%; Score 1052; DB 2; Length 704;
Best Local Similarity 40.1%; Pred. No. 5.7e-64;
Matches 226; Conservative 85; Mismatches 177; Indels 76; Gaps 8;

Qy 2 ASSILKVVVSHQSCRSRSPKRDQREAGSSDLSSRODAENAEAKLGLPGQLVDIACK 61
Db 119 ARNVKLKLVKLEKSSQISIQARE--SEAPEDTQPKQ-----EKLVTYSSDFVDEALS 171
Qy 62 VCOAYLGOLEHEDIDTSADAVEDLTEAWEDLTQOYYSILVHGDAPISNRYFSQCA 121
Db 172 ACQEHLLSSIAHKDIDKDPNSPLVPSDDWSQFLQRYQYQIVHEGAELRYLVEVQRCEDIL 231
RT "High-efficiency full-length cDNA cloning.";

QY 122 NRITSVNPQTDIDGLRNIIWIKPAKSRGRDVIYCMRVEEILELAADHPRLSDNRKVVQ 181
DB 232 QQLQNVVPLDMEGDRNIWIKVKGAKSRGRGICMNRKLDMLKLVDCNPMKMGKQWVQ 291
QY 182 KYIETPLLCIDTKFDIQRWFLVTDWNPILTFWFKESYLRFPSTQRFSLDKLDSAIHLCNNA 241
DB 292 KYIERPELLFGTKFDLRQWFLVTDWNPILTFWFKESYLRFPSTQRFSLDKLDSAIHLCNNA 351
QY 242 VQKYLKNDVGRSPLLPAHNNWTSRTRFOYLQGRGAVGVSIVYPSMKKAIHAAMKVAQD 301
DB 352 IQHLEASCHRRHPLPDPNNWSSQRFQALHQLQVDFADPAKSSVIVPGKAAVHALQTSQD 411
QY 302 HVEPRKNSFELYGADFVLGRDPRFPLWIEINSSPTHPSTPTVTAQLCAQVQEDTIKAV-- 359
DB 412 NVQCRKASFELYGADFVFGEDFQPMWIEINASPTMAPSTAVTARLCAGVQADTLRVVIDR 471
QY 360 --DRSCDIGNFELLWQPVVPPFSGSDLCVAGSVRRARQVLPVCLNKASASLLDAQ 417
DB 472 RLDRSCDITGAFELIYKQPAVEVPQYVGIKLLVEGSTIKK----- 510
QY 418 PLKARGFSAMPDPAQPPSPALORDLGLKEEKLPLALLAPLRGAESGGAQPTRTKAA 477
DB 511 -----PVPVGHRTGVRSS-----LPHLLTQOGSGESKDSGSPTH-RSA 548
QY 478 GKVELPACPRHVDSPQNTGVPVQAP-----AKSWDPNQLNAHPLEPVL-- 523
DB 549 SRKNARAESLEHTEKPEPAVASVSGKGAFFHPSLSKAWLPSPRVHRPQGRVLRQ 608
QY 524 -----GLKTAEGALRPPPGOK 539
DB 609 HDQLVGSKALSTTGKALMTLPTAK 632

RESULT 4

Q9H876 PRELIMINARY; PRT; 744 AA.
AC Q9H876;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ13898.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid gland;
RX PubMed=1470239; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai T., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamiyama K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musahino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Okamoto S.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,

RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizuhina-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
cDNAs";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK023960; BAB14741.1; --
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0008464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_tyf_lygase.
DR Pfam; PF03133; TTB; 1.
KW Ligase.
FT NON TER 744 744
SQ SEQUENCE 744 AA; 84683 MW; DF661753E4AFF0DF CRC64;
Query Match 36.6%; Score 1043.5; DB 2; Length 744;
Best Local Similarity 40.5%; Pred. No. 2.4e-63;
Matches 234; Conservative 85; Mismatches 192; Indels 67; Gaps 12;
QY 2 ASSILKVVVSHQSCSRSSRSPKQDREAEAGSSDLSSRQDAENAEAKLRLPGQLVDIACK 61
DB 151 ARNVKLVVYKSE-----WKSYPIQAVEEASGDKQPKQKPNVL-----VSPFVDEALC 201
QY 62 VQOAYLQGLQHEHEDIDTSADAVEDLTEAEWEDLTQOYYSILVHGDAFTSNRNYFSQCOALL 121
DB 202 ACEEYLSNLAHMDIDKOLEAPLYLTPGWSLFLQRYQVQVHGAELRLHLDTVQRCEDIL 261
QY 122 NRITSVNPQTDIDGLRNIIWIKPAKSRGRDVIYCMRVEEILELAADHPRLSDNRKVVQ 181
DB 262 QQLQNVVPLDMEGDRNIWIKVKGAKSRGRGICMNRKLDMLKLVDCNPMKMGKQWVQ 321
QY 182 KYIETPLLCIDTKFDIQRWFLVTDWNPILTFWFKESYLRFPSTQRFSLDKLDSAIHLCNNA 241
DB 322 KYIERPELLFGTKFDLRQWFLVTDWNPILTFWFKESYLRFPSTQRFSLDKLDSAIHLCNNA 381
QY 242 VQKYLKNDVGRSPLLPAHNNWTSRTRFOYLQGRGAVGVSIVYPSMKKAIHAAMKVAQD 301
DB 382 IQHLEASCHRRHPLPDPNNWSSQRFQALHQLQVDFADPAKSSVIVPGKAAVHALQTSQD 441
QY 302 HVEPRKNSFELYGADFVLGRDPRFPLWIEINSSPTHPSTPTVTAQLCAQVQEDTIKAV-- 359
DB 442 TVQCRKASFELYGADFVFGEDFQPMWIEINASPTMAPSTAVTARLCAGVQADTLRVVIDR 501
QY 360 --DRSCDIGNFELLWQPVVPPFSGSDLCVAGSVRRARQVLPVCLNKASASLLDAQ 417
DB 502 RLDRSCDITGAFELIYKQPAVEVPQYVGIKLLVEGSTIKK-----PMAVCHRRMGVRPAV 555
QY 418 PLKARGPS-----AMPDPAQPPSPALORDLGL-----KEEKG 450
DB 556 PLLTQOGSGEARHHPFSLHTKQA-LPSPHVLRHQVLRQHSKLVGKTKALSTTGKALRT 614
QY 451 LPILA-----LLAPLRGAESGGAQPTRTKA-----AGKVELPACPRHVDSP--AP-NTG 498
DB 615 LPTAKVFISLPNLDKFKVAPSILKPKKAPALLCLRGQPLEVPCCLCPKSEQFLAPVGRS 674
QY 499 VPVQAPKASWDPNQLNAHPLEPVLRLKTAEGALRPPP 536
DB 675 RPKANSRPDCDKPRAEACPMKRL-----SPLKPLP 704
RESULT 5
Q8NDN8 PRELIMINARY; PRT; 434 AA.
ID Q8NDN8
AC Q8NDN8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein DKFZp586B0320.
GN Name=DKFZp586B0320;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL613393; CAD18794.1; -
DR GO: GO:0004835; P:tubulin-tyrosine ligase activity; IEA.
DR GO: GO:0006464; P:protein modification; IEA.
DR InterPro: IPR004344; Tub_tyr_lygase.
DR Pfam: PF03133; TTL; 1.
KW Hypothetical protein.
SQ SEQUENCE 434 AA; 987956CA08651CAL CRC64;

Query Match 34.4%; Score 981.5; DB 2; Length 434;
Best Local Similarity 46.7%; Pred. No. 2.2e-59;
Matches 194; Conservative 64; Mismatches 106; Indels 51; Gaps 5;

QY 74 DIDTSADAVEDLTEAEWEDLTQOYSLVHGDFAFISNSRNFVSCQALLNRITSVNPQTDI 133
DB 2 DIDKDLAPLYLTPEGWSLFQRYQVVGEGALRHLDTVQRCEDILQQLQAVVPQIDM 61

QY 134 DGLRNIIKPAKSRGRDVCMDRVEEILEAAADHPLSRDKNKVVQKVIETPLLCIDT 193
DB 62 EGRNIIWIVPGAKSRGRGIMCHDHEMLKLVNGNPFVWMDGKVVQKVIETPLLCIDT 121

QY 194 KFDIROWFLVDNPLTIWFKSYLRFSTQSFSLDKLDSAIHLCNNAVQKLYKNDVGRS 253
DB 122 KFDLRQWFLVDNPLTIWFKSYLRFSTQSFSLDKLDSAIHLCNNAVQKLYKNDVGRS 181

QY 254 PLPAHNMWTSRFOEVLQQRGAVGWSVIYSMKKAIHAHAKVAQDHVPEPRKNSFELY 313
DB 182 PLPPDNNMSSQRFQALHOGMAGNAPNSTIIVPGMDAVIHALQTSQDTVQCRKASFELY 241

QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAV-DRSCDIGNFE 369
DB 242 GADFVGEDFQPLWIEINASPTWAPSTAVTARLCAGVQADTLRVVIDRMLDNCDTGAFE 301

QY 370 LLWRQPVVEPPFSGDGLCVAGVSRRARRQVLVPCNLKASALLDAQLPKRGPSAMPD 429
DB 302 LIYKQPAVEVQVGIKLLVEGFTIK-PMAM-- 332

QY 430 PAQGPSPALQDLGLKEEKGPLALLAPLGRGAESGGAQPTKKAQKVELPA 484
DB 333 -----CHRRMGVR-----PAVPLLTQRGSGEARHHPFSLHTRK-----QLPS 369

RESULT 6
Q6AWA3 PRELIMINARY; PRT; 352 AA.
AC Q6AWA3;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein DKFZp686D076.
GN Name=DKFZp686D076;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA The German cDNA Consortium;
RA Ansoorge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA Mewes H.W., Weill B., Amid C., Oangner A., Pobo G., Han M., Wiemann S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: BX648175; CAH10554.1; -
DR GO: GO:0004835; F:tubulin-tyrosine ligase activity; IEA.
DR GO: GO:0006464; P:protein modification; IEA.
DR InterPro: IPR004344; Tub_tyr_lygase.
DR Pfam: PF03133; TTL; 1.

KW Hypothetical protein.
SQ SEQUENCE 352 AA; 40257 MW; 49FD8E9CAE1CB20D CRC64;

Query Match 33.5%; Score 957.5; DB 2; Length 352;
Best Local Similarity 54.2%; Pred. No. 7.4e-58;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLTEAEWEDLTQOYSLVHGDFAFISNSRNFVSCQALLNRITSVNPQTDI 133
DB 2 DIDKDLAPLYLTPEGWSLFQRYQVVGEGALRHLDTVQRCEDILQQLQAVVPQIDM 61

QY 134 DGLRNIIKPAKSRGRDVCMDRVEEILEAAADHPLSRDKNKVVQKVIETPLLCIDT 193
DB 62 EGRNIIWIVPGAKSRGRGIMCHDHEMLKLVNGNPFVWMDGKVVQKVIETPLLCIDT 121

QY 194 KFDIROWFLVDNPLTIWFKSYLRFSTQSFSLDKLDSAIHLCNNAVQKLYKNDVGRS 253
DB 122 KFDLRQWFLVDNPLTIWFKSYLRFSTQSFSLDKLDSAIHLCNNAVQKLYKNDVGRS 181

QY 254 PLPAHNMWTSRFOEVLQQRGAVGWSVIYSMKKAIHAHAKVAQDHVPEPRKNSFELY 313
DB 182 PLPPDNNMSSQRFQALHOGMAGNAPNSTIIVPGMDAVIHALQTSQDTVQCRKASFELY 241

QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAV-DRSCDIGNFE 369
DB 242 GADFVGEDFQPLWIEINASPTWAPSTAVTARLCAGVQADTLRVVIDRMLDNCDTGAFE 301

QY 370 LLWRQPVVEPPFSGDGLCVAGVSRRARRQVLVPCNLKASALLDAQLPKRGPSAMPD 429
DB 302 LIYKQPAVEVQVGIKLLVEGFTIK-PMAM-- 332

RESULT 7
TTL3 HUMAN STANDARD; PRT; 352 AA.
ID TTL3 HUMAN
AC Q9Y4R7; Q9U199;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Tubulin tyrosine ligase-like protein 3 (HOTT1) (HQP0207).
GN Name=TTL3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 61-352 FROM N.A.
RC TISSUE=Fetal liver;
RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,
RA Liu M., He F.;
RT "Functional prediction of the coding sequences of 50 new genes deduced
RT by analysis of cDNA clones from human fetal liver."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 TTL domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL096725; CAB46375.1; -
DR EMBL: AF078842; AAF23353.1; -
DR PIR: T12515; T12515.
DR Genew: HGNC:24483; TTL3.
DR InterPro: IPR004344; Tub_tyr_lygase.


```
QY 65 AYLGQLEHEDIDTSADAVEDLTAEWEDLTQOYSLVHGD---AFISNSR-----NYFS 115
Db 238 EYLDYCLHNDIDIEDT--KWDHWDVFTHYLLTHEDNRIOQLKBEEDADAIEHYLA 295
QY 116 QCQALLNRITSVNPQTDIDGLRNIWIKPAKSGRDIQVMDRVEITELEAAADHPLSRD 175
Db 296 EAKSVLEQIKSHWFOYALDGLYNIWIKVPGKRCRGHLMNNIKQI--IAMVNPPIVSK 353
QY 176 NKWVQKIETPLICTKFDIRQWFLVTDWNPITWIFYKESYLRFTSTORFSLDKLDSAI 235
Db 354 TRVIOKIERPLIHTKFDIRQWFMITSVQPLNITWIFYKESYLRFTSSQOYNLMNVHESV 413
QY 236 HLCNNAVOKYLNKDVGSPLPAHNMWTSRFEYLQROGRGAVGWSVIYPSMKKAIAHA 295
Db 414 HLTHAIQKYNHNAV-RDEBLPHENMWDCHTFQAYLRFQIDKYEMWSRIYPGQKAIIGS 472
QY 296 MKVAQDHVPRKNSFELYGADFVLGRDPRFWLWTEINSSPTMHPSTPVTALCAQVOEDTI 355
Db 473 LLACQDMDRRPNTFELYGADFMTEDFYFWLWTEINSSPOLASTSTVTLRCLQCQVEDTI 532
QY 356 KVAVRDSCD----IGNPELLWRQVPEPPFPFSGSDLCVAG 391
Db 533 RVVIDRRTDSNAPTSGFELYKQVIPKTPAYMGLNLQLRG 572
RESULT 10
Q9VM91 PRELIMINARY; PRT; 992 AA.
AC Q9VM91,
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG11323-PA.
OS ORFNames=CG11323;
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abrial J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodek A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush K., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
```

```
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE=22426065; PubMed=12537568;
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Tabor P.S., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
SEQUENCE FROM N.A.
RP MEDLINE=22426070; PubMed=12537573;
RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
SEQUENCE FROM N.A.
RP MEDLINE=22426089; PubMed=12537572;
RX Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whittier E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
SEQUENCE FROM N.A.
RP FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
SEQUENCE FROM N.A.
RP FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003614; AAF52432.1; -.
DR FlyBase; FBgn0031854; CG11323.
DR GO; GO:0004835; P:ubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_ty_lygase.
DR Pfam; PF03133; Ttu; 1.
SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;
Query Match 25.3%; Score 722.5; DB 2; Length 992;
Best Local Similarity 33.3%; Pred. No. 4.8e-41;
Matches 175; Conservative 86; Mismatches 198; Indels 67; Gaps 12;
QY 51 LFGQLVDIACKVQAYLGQLEHEDIDTSADAVED---LTEAEWEDLTQOYSLVHGDFAFI 107
Db 326 IPIPSAIDFAFKRLVEYIDSCQNDID----PEDPPKIWEHDWDAFLFQHQQLVNEGDRI 380
QY 108 SNS-----RNYFSQOALLNRITSVNPQTDIDGLRNIWIKPAKSGRDIQVMDRVEEI 162
Db 381 QHDGGQRLPEFPMVKSLSLVDRMKVHPQYSLDGYQNMWIKVPAKSGRGIILMDNLKKI 440
QY 163 LELAAADHPLSRDNKVVQKVTETPLICTKFDIRQWFLVTDWNPITWIFYKESYLRFS 222
Db 441 --LGVNLSIAKSKRSYIVQVQYIERPLIFQTKDIRQWFLITNTQPLVWVWFYRESYLRFS 498
```



```
Db 353 MVQSTDLVLSNAKFKVSEKVMAYEARLAGLYDQIESLRDPYRDWGSRLNWLKPGYQSR 412
Qy 150 GRDIVCMRVEEIELEAAADHPLSRDNKVVQKVIETPLLICDTKFDIRQWFL--VTDWN 207
Db 413 GIGIVIRSSLDLILQWTSN-----NONKKYIVQKVIETPLLIYRTKFDIRQVMLLTITD-T 467
Qy 208 PLTWFKESYLFRESTQFSLDKLSDIAHLCCNNAVQKLVKNDVGRSPLPAHNNWTSTRF 267
Db 468 KVSITWTRDCVLRFSQSEFTMDLRESIHLTNNSVQKRYKNTRDRLPKNNMWSLDQF 527
Qy 268 QEYLQROGRGAVGMSVTIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFVLGRDFRPL 327
Db 528 KNYLRINGAPDGSWSKTYNGFKQNLVAVVMAISLDETELLQNAFELYGDFMLDHYNPIL 587
Qy 328 IEINSSPTMHPSTVTAQLCAQVEDIRKAVD-----RSCDIGNFELLWRQPVVPPPPFS 383
Db 588 IEINSTPDLSPSTEITARICFVVLKDCIRVVVDLPKNTAATGLFELAFE----- 637
Qy 384 GSDLCVAGSVRRARRQVLVPCVNLKASALLDAQPLKARGPSA-----MPDPAQGPSPSA 438
Db 638 -----VNYISINK-----ADGPLELNGKQMTLFENMPMRNSPTRL 675
Qy 439 LORDLGLKEEGKPLALLAPLRGAESGAAPTRTKAAGKVELPACPCRHVDSPAQNTG 498
Db 676 LRKIL-----NNVKTSTTKVE-----KVVERAPACNVK 703
Qy 499 VPVAQPAKS--WDPN 511
Db 704 NPTAKITKKVWYKN 718

RESULT: 12
Q6ZU95 PRELIMINARY; PRT; 281 AA.
AC Q6ZU95;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein FUJ43887.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK125875; BAC6331.1; -.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 281 AA; 30962 MW; 3F4C4339709E92D0 CRC64;

Query Match 18.4%; Score 524.5; DB 2; Length 281;
Best Local Similarity 40.0%; Pred. No. 4e-28;
Matches 122; Conservative 34; Mismatches 76; Indels 73; Gaps 9;

Qy 233 SAHLCCNNAVQKLVKNDVGRSPLPAHNNWTSTRFQYLQROGRGAVGMSVIVPSMKKAI 292
Db 17 SSVHLCCNNSIQKLENSCHRRPLPPDNNWSSQRFQHLQEMGAPNAWSTIIIVPGMKDAV 76
Qy 293 AHAMKAQDHVEPRKNSFELYGADFVLGRDFRPMWLEINSPTMHPSTVTAQLCAQVQE 352
Db 77 IHALQTSQDTVQCRKASFELYGADFVGFQDFQPLWLEINASPTMAPSTAVTARLCAGVQA 136
Qy 353 DTKIAV-----DRSCDIGNFELLWRQPVVPPPPSGSDLCVAGSVRRARRQVLVPCVNLK 408
```

```
Db 137 DTRVVIDRDLRNCDTGAFELIYKQPAVEVPOYVGRLLVEGFTIKK----- 184
Qy 409 ASASLLDAQPLKARGSPAMPDPAQG--PPSPAL-QRDLGLKEEKG--LPLALLAPLRGAA 463
Db 185 -----PWAMCHRRMGVRPAVPLLTQRGSGAEVSGSLRKLPKVAQLR-RG 228
Qy 464 ESGGAQPTTKAAGKVELPAC-----PCRHVDSOAPNTGV 499
Db 229 TAGMQTQFVTISPA-STFPRSCLLPMYDTRARSSDDSTASWMLRPCR----- 276
Qy 500 PVAQP 504
Db 277 PQARP 281

RESULT 13
Q96GG8 PRELIMINARY; PRT; 101 AA.
AC Q96GG8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TTL3 protein.
GN Name=TTL3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myer R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009479; AAH09479.1; -.
DR GO; GO:0004835; F.tubulin-tyrosine ligase activity; IEA.
DR GO; GO:0006464; P.protein modification; IEA.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 101 AA; 12054 MW; 8EACD6C1131E0DF9 CRC64;

Query Match 13.5%; Score 385; DB 2; Length 101;
Best Local Similarity 68.0%; Pred. No. 4.3e-19;
Matches 68; Conservative 15; Mismatches 17; Indels 0; Gaps 0;

Qy 133 IDGLRWIWIKPAKSRGRDIVCMRVEEIELEAAADHPLSRDNKVVQKVIETPLLICD 192
Db 1 MEGDRNIWIKVPAKSRGRGIMCWDHLEMLKLVNGNVPVWKVQKVIETPLIFG 60
```


Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK015740; BAB29950.1; -
DR MGD; MGI:1913589; 1700016P23Rik.
DR GO; GO:0004835; F:tubulin-cytosine ligase activity; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR004344; Tub_lyr_lygase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 464 AA; 53913 MW; 09FB48E438F433A4 CRC64;

Query Match 12.1%; Score 345; DB 2; Length 464;
Best Local Similarity 29.4%; Pred. No. 1.9e-15;
Matches 86; Conservative 52; Mismatches 95; Indels 60; Gaps 9;

Qy 140 WIIKPAKSRGRDIVCMRVEEILE-----LAAADHPLSRD 175
Db 146 WIMKPVARSQKGIFLLRLKDIMWRKGTSGKKPTGVETOPARANNPNSGSHDTRSD 205

Qy 176 NK-----WYVQYIETPLLCIDTFDIQWFLVTDWNPITWIFYKESYLRFSRFS 228
Db 206 QKDDLPVENYVAQRYVENPYLGGKFDLRVYVLMVSIPLRAMLYRDGPFARFNTFTL 265

Qy 229 DKLDS-AIHLNNAVOK-----YLNVDVGRSPLLPAHNMWTSTRFQSYL-ORQGRGAVWG 281
Db 266 NSIDDHYVHLTNVAVQKTSPTYHLKGC-----KMWLQRFQYLAASKHGFKAV-- 313

Qy 282 SVIYPSMKKAIAHAMKVAQDHVEPRKNSFELYGADFLVGRDPRPMLIEINSSPTMHSTP 341
Db 314 ETLFSDMDNFIKLSQSVQKVIISDKHCFELYGYDILIDQDLKPWLLVNASPSLTASSQ 373

Qy 342 VTAQLCAQVQEDTIKVA-----VDRSCDIGNFELLWRQPVVE----PPPFSG 384
Db 374 EDYELKTCLEDTLHVVDMEARLTGKEKRVGFDLMNDGDPVSRDGDGSDLSG 426

Search completed: September 17, 2005, 09:10:33
Job time : 181 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2005, 08:51:23 ; Search time 172 Seconds
(without alignments)
1216.496 Million cell updates/sec

Title: US-10-615-659-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCRSRR.....LRGLKTAEGALRPPPGKGKS 541

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_16Dec04:*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2854	100.0	541	8 ADJ93358	Adj93358 Human BGS
2	2843	99.6	541	8 ADJ93365	Adj93365 Human tub
3	1560	54.7	293	8 ADJ93366	Adj93366 Human BGS
4	1299	45.5	242	8 ADJ93360	Adj93360 Human BGS
5	1043.5	36.6	744	4 AAB94796	Aab94796 Human pro
6	1013	35.5	488	5 AAU74334	Aau74334 Human Cyt
7	985.5	34.9	399	6 ABU11512	Abu11512 Human MDD
8	969.5	34.0	362	3 AAB58909	Aab58909 Breast an
9	956.5	33.5	352	3 AAB43005	Aab43005 Human ORF
10	956.5	33.5	352	4 AAM39450	Aam39450 Human pol
11	956.5	33.5	352	8 ADJ93457	Adj93457 Human HOT
12	956.5	33.5	352	8 ABM80420	Abm80420 Tumour-as
13	944.5	33.1	326	7 ADM05524	Adm05524 Human pro
14	885.5	30.3	292	8 ADJ93455	Adj93455 Human HOT
15	722.5	25.3	992	4 ABB65645	Abb65645 Drosophil
16	715	25.1	432	6 ABU92048	Abu92048 Human pro
17	536.5	18.8	719	4 ABB65541	Abb65541 Drosophil
18	524.5	18.4	281	8 ADQ66614	Adq66614 Novel hum
19	519	18.2	362	4 AAM41236	Aam41236 Human pol
20	375.5	13.2	566	4 ABG05971	Abg05971 Novel hum
21	336	11.8	160	8 ADH45424	Adh45424 Human mol
22	330	11.6	347	6 ABU00150	Abu00150 Human nov
23	328	11.5	496	4 ABB64074	Abb64074 Drosophil
24	311	10.9	673	8 ADQ66481	Adq66481 Novel hum
25	303.5	10.6	1281	4 AAM39105	Aam39105 Human pol

26	303.5	10.6	1304	4 AAM40891	Aam40891 Human pol
27	299.5	10.5	423	5 ABP43750	Abp43750 Ligase TT
28	299.5	10.5	423	8 ADJ93458	Adj93458 Human tub
29	299	10.5	393	7 ADC31411	Adc31411 Human nov
30	297.5	10.4	827	4 ABB60840	Abb60840 Drosophil
31	291	10.2	989	4 ABB60296	Abb60296 Drosophil
32	282.5	9.9	487	4 ABB59096	Abb59096 Drosophil
33	280.5	9.8	439	7 ADB65504	Adb65504 Human pro
34	278	9.7	321	7 ADC33066	Adc33066 Human nov
35	274	9.6	49	8 ADJ93373	Adj93373 Human BGS
36	271.5	9.5	519	4 AAM79305	Aam79305 Human pro
37	271.5	9.5	524	4 ABB22915	Abb22915 Protein #
38	271.5	9.5	592	7 ADB65217	Adb65217 Human pro
39	270	9.5	917	4 ABB64837	Abb64837 Drosophil
40	259	9.1	379	8 ADJ93456	Adj93456 Pig tubul
41	255.5	9.0	1138	8 ADJ71954	Adj71954 Human PMM
42	254.5	8.9	377	4 AAM79068	Aam79068 Human pro
43	254.5	8.9	377	5 ABB83472	Abb83472 Human cyt
44	254.5	8.9	377	7 ADE47756	Ade47756 Human NOV
45	254.5	8.9	377	8 ADJ79026	Adj79026 Human NOV

ALIGNMENTS

RESULT 1

ADJ93358
ID ADJ93358 standard; protein; 541 AA.

XX AC ADJ93358;

DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 protein sequence SeqID2.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-gen; gastrointestinal-gen;
KW neuroprotective; endocrine-gen; antiinflammatory; anabolic; hypertensive;
KW osteoprotegic; nontropic; antiparkinsonian; antiarthritic; antiaesthetic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human.

Homo sapiens.

WO2004005487-A2.

PD 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

PR 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

DR N-PSDB; ADJ93357.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
disorders.

PS Claim 5; SEQ ID NO 2; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the human
CC BGS-42 protein of the invention.

XX Sequence 541 AA;

Query Match 100.0%; Score 2854; DB 8; Length 541;

Best Local Similarity 100.0%; Pred. No. 3.6e-264;

Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASSILKVVVSHOSCRSRKPRDREAGSDLSRQDAENAEAKLRGLPQLVDIAC 60
Db 1 MASSILKVVVSHOSCRSRKPRDREAGSDLSRQDAENAEAKLRGLPQLVDIAC 60
QY 61 KVCQAYLGQLEHEDIDTSADAVEDLTAEWEDLTQQYYSILVHGDAFISNRYFSQCQAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTAEWEDLTQQYYSILVHGDAFISNRYFSQCQAL 120
QY 121 LNRITSVNPQTDIDGLRNITWIKPAKSGRGRDTCVMDRVEEILELAADHPISRDNKVV 180
Db 121 LNRITSVNPQTDIDGLRNITWIKPAKSGRGRDTCVMDRVEEILELAADHPISRDNKVV 180
QY 181 OKYIETPLICTDKFDIROWFLVTDNPLTIWFYKESYLRFSTQFESLKLDSAIHLN 240
Db 181 OKYIETPLICTDKFDIROWFLVTDNPLTIWFYKESYLRFSTQFESLKLDSAIHLN 240
QY 241 AVQKYLKNDVGRSPLLPAAHNMWTSRFBQYLQQRGAVGWSVIYPSMKKAIAHAMKVAQ 300
Db 241 AVQKYLKNDVGRSPLLPAAHNMWTSRFBQYLQQRGAVGWSVIYPSMKKAIAHAMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDRPRLWLEINSSPTMHPSTPVAQLCAQVQEDTIKAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDRPRLWLEINSSPTMHPSTPVAQLCAQVQEDTIKAVD 360
QY 361 RSCDIGNFELLWQPVVEPPPSGSDLCVAGSVRRARQVLPVNCNLKASASLLDQPLK 420
Db 361 RSCDIGNFELLWQPVVEPPPSGSDLCVAGSVRRARQVLPVNCNLKASASLLDQPLK 420
QY 421 ARGPSAMPDPAQPPSPALQDRLGLKEEGLPLALLAPLRGAESGGAQPTRTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQDRLGLKEEGLPLALLAPLRGAESGGAQPTRTKAAGKV 480
QY 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGTATGALRPPPGGK 540
Db 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGTATGALRPPPGGK 540
QY 541 S 541
Db 541 S 541

RESULT 2

ADJ93365

ID ADJ93365 standard; protein; 541 AA.

XX ADJ93365;

DT 06-MAY-2004 (first entry)

DE Human tubulin tyrosine ligase protein consensus sequence SeqID13.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;

KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;

KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;

KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;

KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;

KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;

KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;

KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;

KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;

KW brain cancer; liver cancer; proliferative condition; testis; lung;

KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;

KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;

KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;

KW sepsis; acne; Sjogren's disease; scleroderma; human.

OS Homo sapiens.

OS Synthetic.

PN WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

DR N-PSDB; ADJ93364.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,

PT useful for preventing, treating or ameliorating a medical condition, e.g.

PT aberrant cellular proliferation, reproductive disorders or testicular

PT disorders.

XX Example 4; SEQ ID NO 13; 343pp; English.

PS This invention relates to a novel testis-specific tubulin tyrosine-ligase

CC -like polypeptide, designated the BGS-42 polypeptide. The invention may

CC be useful for the development of compounds with a cytostatic, respiratory

CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,

CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,

CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,

CC immunosuppressive, antiseborrheic or dermatological activity acting as

CC tyrosine ligase modulators. In addition, the disclosed sequences may be

CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be

CC used for diagnosing a pathological condition or a susceptibility to a

CC pathological condition in a subject, and for preventing, treating or

CC ameliorating a medical condition, such as a disorder related to aberrant

CC tubulin ligase activity, a disorder related to aberrant tubulin-

CC carboxypeptidase activity, aberrant cellular proliferation, reproductive

CC disorders, testicular disorders, testicular cancer, pulmonary disorders,

CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,

CC neural disorders, brain cancer, liver cancer, or proliferative condition

CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42

CC polypeptide, polynucleotide, or their modulators are also useful for

CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's

CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS

-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of the tubulin tyrosine ligase protein consensus sequence which was used in the exemplification of the invention.

XX
SQ Sequence 541 AA;

Query Match 99.6%; Score 2843; DB 8; Length 541;

Best Local Similarity 99.6%; Pred. No. 4.1e-263;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy 1 MASSILKVVSHQSCSRSSRSPKPRDREAGSSDLSSQDAENAEAKLGLPGQVLDIAC 60
Db 1 MASSILKVVSHQSCSRSSRSPKPRDREAGSSDLSSQDAENAEAKLGLPGQVLDIAC 60
Qy 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCOAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNSRNYFSQCOAL 120
Qy 121 LNRITSNPQTDIDGLRNIIWIKPAAKSRGRDIVCMRVBEILEAAADHPLSRDNKVV 180
Db 121 LNRITSNPQTDIDGLRNIIWIKPAAKSRGRDIVCMRVBEILEAAADHPLSRDNKVV 180
Qy 181 QKIETPLLICDTKFDIOWFLVTDWNPITTFYKESYLRESTORFSLDKLSAHLN 240
Db 181 QKIETPLLICDTKFDIOWFLVTDWNPITTFYKESYLRESTORFSLDKLSAHLN 240
Qy 241 AVQYKLVNDVGRSPLPAHNMWTSRFOYLQROGRGAVGWSVTPSPMKKAIAHAMKVAQ 300
Db 241 AVQYKLVNDVGRSPLPAHNMWTSRFOYLQROGRGAVGWSVTPSPMKKAIAHAMKVAQ 300
Qy 301 DHVPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
Db 301 DHVPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
Qy 361 RSDIGNPELLWRQPVVEPPFSGSDLCVAGSVRRARRQVLPVNLKASASLLDAQPLK 420
Db 361 RSDIGNPELLWRQPVVEPPFSGSDLCVAGSVRRARRQVLPVNLKASASLLDAQPLK 420
Qy 421 ARGSPAMPDPAQGPSPALQRDGLKKEKGLPLALLAPLRAAGSAGGAQPTRTKAAGKV 480
Db 421 ARGSPAMPDPAQGPSPALQRDGLKKEKGLPLALLAPLRAAGSAGGAQPTRTKAAGKV 480
Qy 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLVRLGLKTAEGALRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNEHPLVRLSLKTAEGALRPPPGKG 540
Qy 541 S 541
Db 541 S 541
```

RESULT 3
ADJ93366

ID ADJ93366 standard; protein; 293 AA.

XX AC ADJ93366;

DT 06-MAY-2004 (first entry)

XX Human BGS-42 protein-related TTL1 domain.

testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytoskeletal; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; anti-inflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antarthritic; antiaesthetic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary cancer; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;

brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; TTL1 domain.

OS Homo sapiens.

XX WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
useful for preventing, treating or ameliorating a medical condition, e.g.
aberrant cellular proliferation, reproductive disorders or testicular
disorders.

Disclosure; SEQ ID NO 14; 343pp; English.

This invention relates to a novel testis-specific tubulin tyrosine-ligase
-like polypeptide, designated the BGS-42 polypeptide. The invention may
be useful for the development of compounds with a cytostatic, respiratory
-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
anti-inflammatory, anabolic, hypertensive, osteopathic, nontropic,
CC antiparkinsonian, antarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the TTL1
CC domain of the human BGS-42 protein of the invention.

XX SQ Sequence 293 AA;

Query Match

Best Local Similarity 54.7%; Score 1560; DB 8; Length 293;

Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 73 EDITSDA VEDL TEAEWEDLTQYYSLVHGDAFISNSRNYFSQCOALLNRITSVNPQTD 132
Db 1 EDITSDA VEDL TEAEWEDLTQYYSLVHGDAFISNSRNYFSQCOALLNRITSVNPQTD 60
Qy 133 IDGLRNIIWIKPAAKSRGRDIVCMRVBEILEAAADHPLSRDNKVVQKIETPLICD 192
Db 61 IDGLRNIIWIKPAAKSRGRDIVCMRVBEILEAAADHPLSRDNKVVQKIETPLICD 120
Qy 193 TKFDIOWFLVTDWNPITTFYKESYLRESTORFSLDKLSAHLNNAVQYKLVNDVGR 252
Db 121 TKFDIOWFLVTDWNPITTFYKESYLRESTORFSLDKLSAHLNNAVQYKLVNDVGR 180
```


PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.

XX Claim 8; SEQ ID NO 15921; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 744 AA;

Query Match 36.6%; Score 1043.5; DB 4; Length 744;

Best Local Similarity 40.5%; Pred. No. 2.8e-90;

Matches 234; Conservative 85; Mismatches 192; Indels 67; Gaps 12;

QY 2 ASSTLKVVSHQSCRSRSPDRDREAGSSDLSSRQDAENAEAKLRLPGQLVDIACK 61
DB 151 ARNVKLIVKSE-----WKSVPTQAVVEEASGDKPKKQKNPVL-----VSPFVDEALC 201
QY 62 VCQAYLQGLEHEDIDTSADAVEDIATEAWEDLTQYYLSLVHGDFAFISNRNYFSQOALL 121
DB 202 ACEEYLSNLAHMDIDKLEAPLYLTPGWSLFLQRYQVQVHGEAELRHLDTPQVRCEDIL 261
QY 122 NRITSVNPQTDIGLRNIWIIPKAASRGDRIVCMRVBEILLEAAADHPLSRDNKVVQ 181
DB 262 QQLQAVVPOIDMEGRNIWIIPKPAKSGRGIMCMCHLEMLKLIVNGNPNVVMKDGKVVQ 321
QY 182 KYTETPLLICDTKFDIRQWFLVDWNPPLTIWFKESYLRFSTQRFSLDKLDSAIHLGNA 241
DB 322 KYIERPLLIIFGTGKPDQLQWFLVDWNPPLTIWFKESYLRFSTQRFSLDKLDSAIHLGNA 381
QY 242 VQKYLKNDVGRSPLLPANHMTSTRFOYLQQRGAVGWSVIYPSMKKAIAHAMKVAQD 301
DB 382 IQHLENSCHRRPLLPNDNWSQQFQAHLEMGAPNAWSTIIVPGMKDAVIHALQTSQD 441
QY 302 HVEPRKNSFELYGADFVGLDRFRPWLLEINSSPTMHSTPTVTAQLCAQVEDTIKAV-- 359
DB 442 TVQCKKASFELYGADFVGEDFQWMLIEINASTPTMAPSTAVTARLCAVGQADTLRVVIDR 501
QY 360 --DRSCDIGNFELLWRQVVEPPFSGSDLCVAGSVRRARQVLPVCNLKASASILLDAQ 417
DB 502 MLDNRCDTGAFELIYKQPAVEPVQYVIRLVEGFTTK-----PMACHRRMGVRPAV 555
QY 418 PLKARGPS-----AMPDPAQPPSPALQDIDLGL-----KEEGK 450
DB 556 PLLTQSGSGEARHHPFSLHTKQ-LPSPHVLHOGQVLRQHSKLVGTKALSTTGKALT 614
QY 451 LPLA-----LIAPLRGAESGAAPTAKA-----AGKVELPACPCRHVDQ--AP-NTG 498
DB 615 LPTAKVFISLPPNLDLDFKVPASILKPRKAPALLCLLRGPQLEVPCCCLPKSEQFLAPVGRS 674
QY 499 VPVAQPAKSWDPNOLNAPLEPVLRLGKLTAEALRPPP 536

Db 675 RPKANSRPDCDKPRAEACPMKRL-----SPLKPLP 704

RESULT 6

AAU74334

ID AAU74334 standard; protein; 488 AA.

XX AAU74334;

XX 12-MAR-2002 (first entry)

XX Human cytoskeleton-associated protein (CYSKP) #5.

XX Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder;
XX cell proliferative disorder; inflammatory disorder; prion disease;
XX vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
XX neurological disorder; cell motility disorder; reproductive disorder;
XX spinal cord disease; central nervous system disorder; mental disorder;
XX gene therapy; cancer.

XX Homo sapiens.

XX WO200185942-A2.

XX 15-NOV-2001.

XX 03-MAY-2001; 2001WO-US014355.

XX 05-MAY-2000; 2000US-0201960P.

XX 08-MAY-2000; 2000US-0202729P.

XX 05-JUN-2000; 2000US-0209705P.

XX 07-JUN-2000; 2000US-0210149P.

XX 21-JUN-2000; 2000US-0213215P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;

XX Alimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;

XX Policky JL;

XX WPI; 2002-062248/08.

XX N-PSDB; AAS99894.

XX New cytoskeleton-associated proteins and polynucleotides, useful for

XX diagnosing, preventing and treating cell proliferative, autoimmune,

XX inflammatory, neurological, cell motility, reproductive and muscle

XX disorders.

XX Claim 1; Page 130-131; 194pp; English.

XX The invention relates to human cytoskeleton-associated polypeptides

XX (CYSKP) and their associated polynucleotide sequences. The sequences are

XX useful in the treatment of disorders associated with overexpression or

XX underexpression of CYSKP in a patient. The disorders include cell

XX proliferative disorders (such as cancer, actinic keratosis,

XX arteriosclerosis, cirrhosis, hepatitis and psoriasis),

XX autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,

XX osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus

XX and anaemia), vesicle trafficking disorders (such as

XX hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),

XX gastrointestinal disorders, prion diseases, neurological disorders (such

XX as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,

XX Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis

XX and other motor neuron disorders), cell motility disorders, reproductive

XX disorders (such as endometriosis and polycystic ovary syndrome), muscle

XX disorders (such as myocardiitis, migraine, hypertension, hypoglycaemia,

XX myocardial infarction, epilepsy and muscular dystrophy), spinal cord

XX diseases, central nervous system disorders (such as Down syndrome and

XX cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

XX Sequences AAU74330-AAU74363 represent human CYSKP of the invention

XX
XX Sequence 488 AA;

XX	Query Match	35.5%; Score 1013; DB 5; Length 488;
PA	Best Local Similarity	45.5%; Pred. No. 1.2e-87;
XX	Matches	214; Conservative 68; Mismatches 138; Indels 50; Gaps 10;
QY	74	DIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNRVFSQOALLNRITSVNQTDI 133
DB	2	DIDKLEAPLYLTPGWSLFQRYQVHVHGAELRLDTQVQCEDILOQLQAVVPQIDM 61
QY	134	DGLNRNWIIPKAASRGDRDVCMDRVEEILEAAADHPLSRDKNRKVVVQKYIETPLLICDT 193
DB	62	EGRNRIWIKFGAKSRGRGIMCMHDEMLKLVNGPNVVKQKVVQKYIERPLLI 121
QY	194	KFDIRQWFLVDNPLTIWFIKESYLRFSTQRFSLDKLDSAIHLNCNNAVKQYKLVNDVGRS 253
DB	122	KFDLRQWFLVDNPLTIWFIKESYLRFSTQRFSLDKLDSAIHLNCNNSIQKLENSCHRR 181
QY	254	PLLPANMWTSTFOEYLQOGRGAVNGSVIYPSMKKATAHANKVAQDHVEPRKNSFELY 313
DB	182	PLLPANMWSQSOFQAHLENGAPNASTIIVPGMKDAVTHALQTSQDVTQCRKASFELY 241
QY	314	GADFVLGRDPRPWLIEINSSPTHPSTPTVTAQLCAOVQEDTIKAVV---DRSCDIGNFE 369
DB	242	GADFVGEDEFPWLIEINASTPTAPSTAVTARLCAGVQATLRVIDMLDRNCDTGAFZ 301
QY	370	LLWRQVPEPPPSGSDLCVAGSVVRRARRQVLPVNCNLKASALLDAQPLKARGPSAMPD 429
DB	302	LIYQKPAVEVPQVIGIRLLVEGTIKK-----PWAMCH 334
QY	430	PAQG--PPSPAL-ORDLGLKEEGLPLALLAPLR--GAESGGAQP---TRTKAAGK-- 479
DB	335	RRMGVRAVPELLTQSGEGKDSGIPTHRSASRKGTCARSLGHSEKPEVSTATTSPGKGK 394
QY	480	----VELPACPCRHV--DSQAPNT--GVPEAQAQAKSWD--PNQLNAHPLEP 520
DB	395	KGKAKATALLVCPNLWEWDAPSTRMGCIFTFTFSSGDRQPHNLRLPLSP 444
RESULT 7		
ABU11512		
ID	ABU11512	standard; protein; 399 AA.
XX	AC	ABU11512;
XX	DT	12-FEB-2003 (first entry)
DE	Human	MDDT polypeptide SEQ ID 459.
XX	KW	MDDT; human; disease detection and treatment molecule polypeptide;
KW		anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-Hiv;
KW		haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
KW		gene therapy; protein replacement therapy; cell proliferative disorder;
KW		cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
KW		anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW		Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW		psoriasis; hepatitis.
OS	Homo sapiens.	
XX	PN	WO20027949-A2.
XX	PD	10-OCT-2002.
XX	PF	27-MAR-2002; 2002WO-US009944.
XX	PR	28-MAR-2001; 2001US-0279619P.
PR		29-MAR-2001; 2001US-0280067P.
PR		29-MAR-2001; 2001US-0280068P.
PR		16-MAY-2001; 2001US-0291280P.
PR		17-MAY-2001; 2001US-0291829P.
PR		17-MAY-2001; 2001US-0291849P.
PR		19-JUN-2001; 2001US-0299428P.
PR		20-JUN-2001; 2001US-0299776P.
PR		20-JUN-2001; 2001US-0300001P.

XX	(INCY-) INCYTE GENOMICS INC.	
PA	Daffo A. Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;	
XX	Daffo GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;	
PI	Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstein EH;	
PI	Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;	
PI	Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;	
XX	WPI: 2003-058431/05.	
DR	N-PSDB; ABX34502.	
XX		
PT	New purified disease detection and treatment molecule proteins and	
PT	polynucleotides, useful for diagnosing, treating or preventing cancers	
PT	(e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis	
PT	or hepatitis.	
XX		
PS	Claim 27; SEQ ID NO 459; 339pp + Sequence Listing; English.	
XX		
CC	This invention describes a novel disease detection and treatment molecule	
CC	polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,	
CC	osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,	
CC	antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides	
CC	and the polypeptides of the invention can be used for gene therapy,	
CC	protein replacement therapy and are useful for treating a variety of	
CC	diseases or conditions. These polypeptides or polynucleotides are	
CC	particularly useful for diagnosing, treating or preventing cell	
CC	proliferative disorders (e.g. cancers including adenocarcinoma,	
CC	leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's	
CC	disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's	
CC	syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or	
CC	hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded	
CC	by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:	
CC	The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences	
XX		
SQ	Sequence 399 AA;	
Query Match		34.9%; Score 995.5; DB 6; Length 399;
Best Local Similarity		53.4%; Pred. No. 4.3e-86;
Matches		187; Conservative 57; Mismatches 101; Indels 5; Gaps 2;
QY	54	QLVDIACKVCOAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQYYSLVHGDAFISNRNY 113
DB	29	EFVDEALCACEEYLSNLAHMDIDKLEAPLYLTPEGMSLFQRYQVHVHGAELRLDTQ 88
QY	114	FSCQALNRITSVNQTDIGLRNIWIKPAASRGDRDVCMDRVEEILEAAADHPLS 173
DB	89	VQRCEDILQQLQAVVPQIDMEGRNIWIKPGAASRGDRDVCMDRVEEILEAAADHPLS 148
QY	174	RNKXVQVQKYIETPLLICDTKFDIRQWFLVDNPLTIWFIKESYLRFSTQRFSLDKLDS 233
DB	149	KDKGKVVQKYIERPLLIETGKFDLRQWFLVDNPLTIWFIKESYLRFSTQRFSLDKLDS 208
QY	234	AIHLNCNNAVKYLVNDVGRSPLLPANMWTSTFOEYLQOGRGAVNGSVIYPSMKKAIA 293
DB	209	SVHLNCNNSIQKLENSCHRRPLLPDNNWSSQRFQAHLENGAPNASTIIVPGMKDAVI 268
QY	294	HANKVAQDHVEPRKNSFELYGADFVLGRDPRPWLIEINSSPTHPSTPTVTAQLCAOVQED 353
DB	269	HALQTSQDVTQCRKASFELYGADFVFGSDFPWLIEINASTPTAPSTAVTARLCAGVQAD 328
QY	354	TIKAVV---DRSCDIGNFELLWRQVPEPPPSGSDLCVAGV-SVRRAR 398
DB	329	TURVVIDRLDRNCDTGAFELIYKQVTTSPASTPRPSCLLPMYSUTRAR 378
RESULT 8		
AAB58909		
ID	AAB58909	standard; protein; 362 AA.
XX		
AC	AAB58909;	

XX 27-MAR-2001 (first entry)

DT Breast and ovarian cancer associated antigen protein sequence SEQ ID 617.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;

XX neutropic; neuroprotective; antiviral; antiallergic; hepatotropic;

XX antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;

XX antibacterial; antifungal; antiparasitic; cardiant; immune disorder;

XX Addison's disease; allergy; autoimmune haemolytic anaemia;

XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;

XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;

XX cardiovascular disorder; wound healing; neurological disease.

OS Homo sapiens.

XX WO200055173-A1.

FN 21-SEP-2000.

PD 08-MAR-2000; 2000WO-US005881.

PF 12-MAR-1999; 99US-0124270P.

PR (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM;

PI WPI; 2000-611515/58.

DR N-PSDB; AAF21812.

DR New human breast and ovarian cancer associated gene sequences and the

XX polypeptides encoded by these genes, useful in the prevention, treatment

PT and diagnosis of cancer, immune disorders, cardiovascular disorders and

PT neurological diseases.

XX Claim 11; Page 1056-1057; 1299pp; English.

PS Sequences AAF21614 - AAF2031 represent DNA sequences encoding human

CC proteins AAB58711 - AAB59128. The DNA and protein sequences are

CC associated with breast and ovarian cancer. Included in the invention are

CC sequences AAF2032 - AAF22040 and AAB59129 which are used in the

CC isolation and characterisation of the DNA and protein sequences of the

CC invention. The breast and ovarian cancer associated DNA, protein, agonist

CC or antagonist sequences exhibit cytostatic; immunosuppressive; neutropic;

CC neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;

CC antiinflammatory; antitumor; vulnery; anticonvulsant; antibacterial;

CC antifungal; antiparasitic and cardiant activity. The polynucleotide and

CC protein sequences are used in the diagnosis of cancer, particularly

CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists

CC and agonists may also be used in the diagnosis, prevention and treatment

CC of immune disorders e.g. Addison's disease, allergies, autoimmune

CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;

CC cardiovascular disorders such as myocardial ischaemia; wound healing;

CC neurological diseases such as cerebral anoxia and epilepsy; and

CC infectious diseases

XX SQ Sequence 362 AA;

Query Match 34.0%; Score 969.5; DB 3; Length 362;

Best Local Similarity 52.0%; Pred. No. 1.1e-83;

Matches 183; Conservative 58; Mismatches 92; Indels 19; Gaps 3;

Qy 49 RGLPQLVDIACKVQAVLGLEHEDITSADAVEDLTAESWEDLTQOYSLVHGDAFIS 108

Db 9 RGVYQ-----LGPHGHRQ---GLEAPLYLTPEGWLSFLGRYQVYVHGELR 53

Qy 109 NSRNYFSQCALLNRITSVNPQTDIDGLRNIIWKPAKSRGRDIVCMRVEEILELAAA 168

Db 54 HLDFTQVQRCELDLQQLQAVVQIDMEGDRNIWIKPGAKSRGRGIMCMHLEMLKLVNG 113

Qy 169 DHPULSRDNKVVQVKYIETPLLICDTKFDIQRQFLVTDNPNLTIWPKYKESYLRFTQPSL 228

Db 114 NPVVMKDGKVVQKIERPLIFGFKDLRQWFLVTDNPLTWVFRDSYLRFTQPSL 173

Qy 229 DKLDSAIHLNNAVQKYLKNDVGRSPLLPAHNMVMTSTRFQYLRQGRGAVGWSVIYPSM 288

Db 174 KNLDSNVHLNNSIQKLENSCHRHPLLPDNNMSSQRFQAHLCQMGAPNAWSTIIVPGM 233

Qy 289 KKAIAHAMKVAQDHVERKNSFELYGADFVLGRDPRKWLTEINSPTMHPSTPTVTAQLCA 348

Db 234 KDAVIHALQTSQDTVQCRKASFELYGADFVFGEDFQPLWLEINASPTMAPSTAVTARLCA 293

Qy 349 QVQSDTIKAV---DRSCDIGNFELLWQPVVPPPPFSGSDLCVAGSVVR 396

Db 294 GVQADTLRVVIDRDLDRNCDTGAFELIYKQPAVEVQYVGIRLLVGEFTIKK 345

RESULT 9

AAB43005

ID AAB43005 standard; protein; 352 AA.

XX AAB43005;

AC 08-FEB-2001 (first entry)

XX Human ORFX ORF2769 polypeptide sequence SEQ ID NO:5538.

DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;

XX vulnery; antiparasitic; antiparkinsonian; neutropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

XX thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US008621.

XX 31-MAR-1999; 99US-0127607P.

PR 02-APR-1999; 99US-0127636P.

PR 05-APR-1999; 99US-0127728P.

PR 30-MAR-2000; 2000US-00540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

DR N-PSDB; AAC77214.

XX Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease.

XX Claim 11; Page 4720-4721; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

CC antiparasitic; antiparkinsonian; neutropic; neuroprotective; osteopathic;

CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;

CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;

CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORF
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
CC
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 3; Length 352;
Best Local Similarity 54.2%; Pred. NO. 1.9e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLTAEWEDLTQOYSLVHGDAFISNRNVSFQCALLNRITSVNPQTDI 133
DB 2 DIDKLEAPLYLTPEGWSLFLQRYQVQVHGAELRLDTQVQRCEDILQQLQAVVPQIDM 61
QY 134 DGLRNWIIPKAASGRDIVCMRDVEEILELAAADHPLSRDNKWWVQKYEIETPLLICDT 193
DB 62 EGDNRNIWVPGAKSRGIMCMCHLEMLKLVNGPNVVMKDGKWWVQKYEIETPLIFGT 121
QY 194 KFDIRQWFLVTDWNPITWFIYKESYLRFSQFSLDKLSAHLKNNVQKYLKNDVGVS 253
DB 122 KFDLRQWFLVTDWNPITWFIYKESYLRFSQFSLDKLSAHLKNNVQKYLKNDVGVS 181
QY 254 PLLPAHNMWTSRFOEYLQRCRGVAGWSVYISMKKAIAHAMKVAQDHVEPKNSFELY 313
DB 182 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDITVQCKASFELY 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAV-DRSCDIGNFE 369
DB 242 GADVFGEDEFQWLIEINASPTMAPSTAVTARLCAGVQADTLRVVDRRLDRNCDTGAFE 301
QY 370 LLWRQPVVEPPFSGSDLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 10
AAM39450
ID AAM39450 standard; protein; 352 AA.
XX
AC AAM39450;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 2595.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

PN 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Zhou P, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Qa;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58606.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2595; 10078pp; English.
PS
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 4; Length 352;
Best Local Similarity 54.2%; Pred. NO. 1.9e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DIDTSADAVEDLTAEWEDLTQOYSLVHGDAFISNRNVSFQCALLNRITSVNPQTDI 133
DB 2 DIDKLEAPLYLTPEGWSLFLQRYQVQVHGAELRLDTQVQRCEDILQQLQAVVPQIDM 61
QY 134 DGLRNWIIPKAASGRDIVCMRDVEEILELAAADHPLSRDNKWWVQKYEIETPLLICDT 193
DB 62 EGDNRNIWVPGAKSRGIMCMCHLEMLKLVNGPNVVMKDGKWWVQKYEIETPLIFGT 121
QY 194 KFDIRQWFLVTDWNPITWFIYKESYLRFSQFSLDKLSAHLKNNVQKYLKNDVGVS 253
DB 122 KFDLRQWFLVTDWNPITWFIYKESYLRFSQFSLDKLSAHLKNNVQKYLKNDVGVS 181
QY 254 PLLPAHNMWTSRFOEYLQRCRGVAGWSVYISMKKAIAHAMKVAQDHVEPKNSFELY 313
DB 182 PLLPPDNMWSQRFOAHLQEMGAPNAWSTIIVPGMKDAVIHALQTSQDITVQCKASFELY 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKAV-DRSCDIGNFE 369
DB 242 GADVFGEDEFQWLIEINASPTMAPSTAVTARLCAGVQADTLRVVDRRLDRNCDTGAFE 301
QY 370 LLWRQPVVEPPFSGSDLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSPASTPRPSCLLPMYSSTRAR 331

RESULT 11
ADJ93457
ID ADJ93457 standard; protein; 352 AA.
XX

ADJ93457;
06-MAY-2004 (first entry)
Human HOTT3 protein sequence SeqID2.
testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory; Gen; gastrointestinal; Gen;
neuroprotective; endocrine; Gen; anti-inflammatory; anabolic; hypertensive;
osteopathic; neotropic; antiparkinsonian; antiarthritic; antiaesthetic;
anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; HOTT3.
Homo sapiens.
WO2004005487-A2.
15-JAN-2004.
09-JUL-2003; 2003WO-US021605.
09-JUL-2002; 2002US-0394725P.
(BRIM) BRISTOL-MYERS SQUIBB CO.
Feder JN, Wu S, Nelson TC;
WPI; 2004-099381/10.
New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, e.g.
useful for preventing, treating or ameliorating a medical condition, e.g.
aberrant cellular proliferation, reproductive disorders or testicular
disorders.
Disclosure; SEQ ID NO 7; 343pp; English.
This invention relates to a novel testis-specific tubulin tyrosine-ligase
-like polypeptide, designated the BGS-42 polypeptide. The invention may
be useful for the development of compounds with a cytosolic, respiratory
-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
anti-inflammatory, anabolic, hypertensive, osteopathic, neotropic,
antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
immunosuppressive, antiseborrheic or dermatological activity acting as
tyrosine ligase modulators. In addition, the disclosed sequences may be
useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
used for diagnosing a pathological condition or a susceptibility to a
pathological condition in a subject, and for preventing, treating or
ameliorating a medical condition, such as a disorder related to aberrant
tubulin ligase activity, a disorder related to aberrant tubulin-
carboxypeptidase activity, aberrant cellular proliferation, reproductive
disorders, testicular disorders, testicular cancer, pulmonary disorders,
lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
neural disorders, brain cancer, liver cancer, or proliferative condition
of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
polypeptide, polynucleotide, or their modulators are also useful for
treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
-42 polypeptide can be used as a preventive agent for immunological
disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
disease or scleroderma. The antibodies may be used to purify, detect and
target the BGS-42 polypeptides. The present sequence is that of the human
HOTT3 protein which is related to the invention. Note: The present
sequence does not appear in the specification but was obtained from
Genbank.

SQ Sequence 352 AA;
Query Match 33.5%; Score 956.5; DB 8; Length 352;
Best Local Similarity 54.2%; Pred. No. 1.9e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;
Qy 74 DIDTSADAVDLTEAWEDLTQQYYSLVHGDFTSNRSNRYFSQCOALLNRTTSVNPQTDI 133
Db 2 DIDKDLAPLYLTPEGWSLFLQRYQVHGEAEURHLDTVQRCEDILQQQAVVPQIDM 61
Qy 134 DGLRNIWIIKPAKSRGRDIVCMRVEEILELAADHPLSRDNKVVVQKYIETPLLICDT 193
Db 62 EGDNRNIWIKPAKSRGRGIMCMHLEMLKLVNGNPMVKDGKVVQKYIERPLIFGT 121
Qy 194 KFDIRQWFLVTDNPLTIWFKESYLFSTQRFSLDKLDSAIHLGNNAVQKYLKNDVGRS 253
Db 122 KFDLRQWFLVTDNPLTIWFKESYLFSTQRFSLDKLDSAIHLGNNSIQHLENSCHRH 181
Qy 254 PLLPAHNMWTSRPOEYLRQGRGAVGVSIVYPSMKKAIAHAMKVAQDHVPRKNSFELY 313
Db 182 PLLPDDNWSQRFOAHLQEMGAPNASTIIVPMKDAVHALQTSQDTVCRKASFELY 241
Qy 314 GADFLGRDPRPWLIEINSSPTMHPSTPTAQLCAQVQEDTIKVAV----DRSCDIGNFE 369
Db 242 GADPFGEDFQPLWIEINASTPAPSTAVTARLCAGVQADTLRVVIDRMLDNCDDTGA 301
Qy 370 LLWRQPVVEPPPPSGSLCVAGV-SVRRAR 398
Db 302 LIYKQPVTTSPASTPRPSCLLPMYSDTRAR 331
RESULT 12
ABM80420
ID ABM80420 standard; protein; 352 AA.
XX AC ABM80420;
XX DT 18-NOV-2004 (first entry)
XX DE Tumour-associated antigenic target (TAT) polypeptide PRO80956, SEQ:1058.
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX OS Homo sapiens.
XX XX WO2004030615-A2.
XX XX 15-APR-2004.
XX XX 29-SEP-2003; 2003WO-US028547.
XX XX 02-OCT-2002; 2002US-0414971P.
XX XX (GSETH) GENENTECH INC.
XX FI Wu TD, Zhang Z, Zhou Y;
XX WPI; 2004-347921/32.
XX DR N-PSDB; ACN37881.
XX XX New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX XX Claim 12; SEQ ID NO 1058; 7273pp; English.
XX XX

CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX
SQ Sequence 352 AA;

Query Match 33.5%; Score 956.5; DB 8; Length 352;
Best Local Similarity 54.2%; Pred. No. 1.9e-82;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY 74 DITSDAVEDLTEAEWEDLTQYYSIVLHGDAFISNRNYSFSCQALLNRITSVNPQTDI 133
DB 2 DIDKDLEAPLYLTPEGWSLFLQRYQVVEGAEHLRLDTQVQCEDILQQLQAVVFOIDM 61
QY 134 DGLRNIIWIKPAKSRGRDIVCMRDVEEILELAADHPLSRDNKWWVQVQYIETPLIICDT 193
DB 62 EGDENIWIWPKGAKSRGRGIMCMHLEMLKLVNGPNVVMKDGKWWVQVQYIERPLIJFT 121
QY 194 KFDIROWFLVTDNPLTIWTFYKESYLRFSQPSLQKLSAHLCHNNAVKYLNKDVGRS 253
DB 122 KFDLQWFLVTDNPLTIWTFYRDSYIRFSTQPSLKNLNSVHLNNSIQKHLENSCHRH 181
QY 254 PLLPAHNMWTSRFOELQQRGAVGWSVIVPSMKKATAHAKVQADHVEPKNSFELY 313
DB 182 PLLPPDNMSSQRFQAHQBMGAPNASTIIVPGMKDAVIHALQTSQDTVQCKKASFE 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAV-DRSCDIGNFE 369
DB 242 GADFVFGEDFQPLWLEINASTPTMAPSTAVTARLCAGVQADTLRVVDRDLRNCDTGAFE 301
QY 370 LLWRQPVPEPPSPGSDLCVAGV-SVRRAR 398
DB 302 LIYKQPVTTSPASTPRPSCLLPMYSIDTRAR 331

RESULT 13
ADM05524
ID ADM05524 standard; protein; 326 AA.
XX
AC ADM05524;

DT 20-MAY-2004 (first entry)
XX
DE Human protein of the invention SEQ ID NO:4209.
XX
KW human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.

XX
PA
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
DR WPI; 2003-723558/69.
DR N-PSDB; ADM03081.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 4209; 305pp; English.

XX The invention relates to a novel human polynucleotide and the encoded
XX polypeptide. A polynucleotide of the invention may have a use in gene
XX therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
XX as a primer for synthesizing the polynucleotide or as a probe for
XX detecting the polynucleotide. The polynucleotides ADM0316-ADM03758 are
XX useful in gene therapy, for developing a diagnostic marker or medicines
XX for regulating their expression and activity, or as a target of gene
XX therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
XX are useful as pharmaceutical agents. The present sequence represents a
XX protein sequence of the invention.

XX Sequence 326 AA;

Query Match 33.1%; Score 944.5; DB 7; Length 326;
Best Local Similarity 54.3%; Pred. No. 2.4e-81;
Matches 176; Conservative 55; Mismatches 86; Indels 7; Gaps 2;

QY 74 DITSDAVEDLTEAEWEDLTQYYSIVLHGDAFISNRNYSFSCQALLNRITSVNPQTDI 133
DB 2 DIDKDLEAPLYLTPEGWSLFLQRYQVVEGAEHLRLDTQVQCEDILQQLQAVVFOIDM 61
QY 134 DGLRNIIWIKPAKSRGRDIVCMRDVEEILELAADHPLSRDNKWWVQVQYIETPLIICDT 193
DB 62 EGDENIWIWPKGAKSRGRGIMCMHLEMLKLVNGPNVVMKDGKWWVQVQYIERPLIJFT 121
QY 194 KFDIROWFLVTDNPLTIWTFYKESYLRFSQPSLQKLSAHLCHNNAVKYLNKDVGRS 253
DB 122 KFDLQWFLVTDNPLTIWTFYRDSYIRFSTQPSLKNLNSVHLNNSIQKHLENSCHRH 181
QY 254 PLLPAHNMWTSRFOELQQRGAVGWSVIVPSMKKATAHAKVQADHVEPKNSFELY 313
DB 182 PLLPPDNMSSQRFQAHQBMGAPNASTIIVPGMKDAVIHALQTSQDTVQCKKASFE 241
QY 314 GADFVLGRDPRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAV-DRSCDIGNFE 369
DB 242 GADFVFGEDFQPLWLEINASTPTMAPSTAVTARLCAGVQADTLRVVDRDLRNCDTGAFE 301
QY 370 LLWRQ---PVPEPPSPGSDLCVA 390
DB 302 LIYKQGPAPNMQVSPERNAPLCPA 325

RESULT 14
ADJ93455
ID ADJ93455 standard; protein; 292 AA.

XX
AC ADJ93455;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human HOTT1 protein sequence SeqID2.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antilasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;

108	SNS-----RNYFSQCALLNRITSVNPTDDIGLRLNTWI1KPAAKSRGRDVCMDRVEEI	162
Qy		
381	QHDGGORLEPMVYKSLSLVDRKMHWPQYSLDGYNM1VKPANKCRGRIILMDNLKKI	440
Db		
163	LELAADHPLSRDKNWVQKY1TETPLLICD7KFDIOWFLVTDWNP1TIWFKYESVLRFS	222
Qy		
441	--LGVNLSIAKSRVYVQKYIERPL11L1FQ7KFDIOWFL1TQPLVWVFYRESVLRFS	498
Db		
223	TQRFSLDKLDLSAIHLGNNAVQKYLKNDVGRSPL1PAHNM7TSTRFBYLRQORGRGAVWS	282
Qy		
499	SOEYSLNSHHESVHLTNTAIQKYTYNG-KRDKRLPSENMMWDCYSFOAYLRQIKGYNNMLE	557
Db		
283	VLYPSMKKA1AHAMKVAODHVEPRKNSFELYGADPVLGRDPRFWL1EINSSPTMHPSTVP	342
Qy		
558	R1FFPGMRKA1VCGMLASQENMDRRNTEFELEGAOFM1CENFYPLW1EINSSPDLGATTSV	617
Db		
343	TALQCAQVOEDTIKVAVDNRSCD-----IGNFELLWRQVPVPPFPFSGSDLCVAGVSV	394
Qy		
618	TARMCPQCLEDVVKVVIDRRDTPKAE1GNFELAYRVQVVPPTPATMGLN1LPVKQKVLQKA	677
Db		
395	-----RRARQVLPVCM1KASALLDAQPLKARGSPAMP-----	428
Qy		
678	NHGGGHGHHYVYQOQRKERSLATSVYRORS1AIHPATS1SRIHRAMP7TFNATEYMEKYMW	737
Db		
429	DPAQGPSPALQORD1GLKEEGLPL1ALLAPURGAASGGAAQPTRTKAAGKVELPACPCK	488
Qy		
738	EPLSSRS--SLCSQLPKQSPSNA1PATP-----SGATSSYLKQAGRSITQLLSAT	789
Db		
489	H-----VDSQAPNTGVVPAQPAKSWDNPQLNAH1PLEPVLRGLK	526
Qy		
790	HKRNTGSLSGBOVOSTALP-PKQRSOCGRPLSTNPVESTKEFKF	834
Db		

Search completed: September 17, 2005, 09:07:30
Job time : 175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 23, 2005, 23:48:14 ; Search time 698.935 Seconds
(without alignments)
8666.982 Million cell updates/sec

Title: US-10-615-659-1_COPY_369_1274
Perfect score: 906
Sequence: 1 gaggacatcacacgtcagc.....tcgagctcctgtgaggcag 906

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
- 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	906	100.0	1838	19	US-10-615-659-1
2	906	100.0	1838	19	US-10-615-659-1
3	906	100.0	3554	19	US-10-615-659-12
4	906	100.0	3554	19	US-10-615-659-12
5	806.8	89.1	1859	19	US-10-615-659-10
6	806.8	89.1	1859	19	US-10-615-659-10
7	762	84.1	3465	19	US-10-615-659-11
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 12, Appl
					Sequence 12, Appl
					Sequence 10, Appl
					Sequence 10, Appl
					Sequence 11, Appl

8	762	84.1	3465	19	US-10-615-659-11	Sequence 11, Appli
9	760.4	83.9	1939	19	US-10-615-659-9	Sequence 9, Appli
10	760.4	83.9	1939	19	US-10-615-659-9	Sequence 9, Appli
11	726	80.1	726	19	US-10-615-659-3	Sequence 3, Appli
12	726	80.1	726	19	US-10-615-659-3	Sequence 3, Appli
c 13	427	47.1	490	16	US-10-029-386-24894	Sequence 24894, A
c 14	427	47.1	101270	20	US-10-723-860-631	Sequence 631, App
15	398.8	44.0	2380	18	US-10-275-595A-39	Sequence 39, Appl
16	398.8	44.0	2553	15	US-10-102-524-1765	Sequence 1765, Ap
17	398.8	44.0	2553	21	US-10-956-157-2016	Sequence 2016, Ap
18	398.8	44.0	2553	22	US-10-756-149-1971	Sequence 1971, Ap
19	398.8	44.0	2848	17	US-10-108-260A-1766	Sequence 1766, Ap
20	398.8	44.0	3001	15	US-10-037-270-494	Sequence 494, App
21	398.8	44.0	3001	17	US-10-117-722-494	Sequence 494, App
22	390.6	43.1	2611	9	US-09-925-298-199	Sequence 199, App
23	390.6	43.1	2611	14	US-10-102-806-199	Sequence 199, App
24	378.4	41.8	1400	21	US-10-956-157-9356	Sequence 9356, Ap
25	378.4	41.8	1673	21	US-10-956-157-4121	Sequence 4121, Ap
26	316.4	34.9	1400	21	US-10-956-157-7251	Sequence 7251, Ap
27	196	21.6	755	13	US-10-027-632-135929	Sequence 135929,
28	196	21.6	755	13	US-10-027-632-135930	Sequence 135930,
29	196	21.6	755	17	US-10-027-632-135929	Sequence 135929,
30	196	21.6	755	17	US-10-027-632-135930	Sequence 135930,
31	190.6	21.0	1958	21	US-10-956-157-4903	Sequence 4903, Ap
32	166	18.3	1728	22	US-10-450-763-5962	Sequence 5962, Ap
33	164.4	18.1	2250	22	US-10-450-763-5187	Sequence 5187, Ap
34	163.4	18.0	418	9	US-09-983-965-2032	Sequence 2032, Ap
c 35	130	14.3	525	16	US-10-029-386-11174	Sequence 11174, A
36	122.6	13.5	436	17	US-10-242-535A-7660	Sequence 7660, Ap
37	122.6	13.5	436	18	US-10-085-783A-7660	Sequence 7660, Ap
38	117	12.9	492	18	US-09-918-995-35253	Sequence 35253, A
39	113.4	12.5	279	18	US-10-424-599-119452	Sequence 119452,
40	92.2	10.2	207	9	US-09-728-445-450	Sequence 450, App
41	92.2	10.2	207	22	US-10-964-549-450	Sequence 450, App
c 42	80	8.8	80	19	US-10-615-659-52	Sequence 52, Appl
c 43	80	8.8	80	19	US-10-615-659-52	Sequence 52, Appl
c 44	58.2	6.4	1648	21	US-10-887-553A-838	Sequence 838, App
45	56.8	6.3	1080	15	US-10-156-761-2625	Sequence 2625, Ap

ALIGNMENTS

RESULT 1
US-10-615-659-1
; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: P0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1

Query Match 100.0%; Score 906; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 1.8e-260;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGGACATCACACGTGAGGAGTCCGTGAGGAGCTACTGAGCGGAGTGGAGGAC 60
|||||


```
Db 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGAC 428
QY 61 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAAT 120
Db 429 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAAT 488
QY 121 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCACGTCGTGTGAACCTTCAGACGGAC 180
Db 489 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCACGTCGTGTGAACCTTCAGACGGAC 548
QY 181 ATTGACGGCTCCGGAAACATCTGGATATTAAGCCCGCCGCAAGTCCCGGGGCCGAGAC 240
Db 549 ATTGACGGCTCCGGAAACATCTGGATATTAAGCCCGCCGCAAGTCCCGGGGCCGAGAC 608
QY 241 ATAGTGTGATGACACGTCGTGGAGGAGATCCTGGAGCTGGCAGCTCGACACCAACCTCTT 300
Db 609 ATAGTGTGATGACACGTCGTGGAGGAGATCCTGGAGCTGGCAGCTCGACACCAACCTCTT 668
QY 301 TCCAGGGAACAAGTGGGTGCTCCAGAAATGATACATCGAGACGCCGCTGCTCATCTGTGAC 360
Db 669 TCCAGGGAACAAGTGGGTGCTCCAGAAATGATACATCGAGACGCCGCTGCTCATCTGTGAC 728
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGGACACCCCTGACCATCTGG 420
Db 729 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGGACACCCCTGACCATCTGG 788
QY 421 TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 480
Db 789 TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 848
QY 481 AGCGCCATCCACTGTGCAACAAACGCGCTCCAGAAATGATCTGAAGAAATGATGTGGGCCG 540
Db 849 AGCGCCATCCACTGTGCAACAAACGCGCTCCAGAAATGATCTGAAGAAATGATGTGGGCCG 908
QY 541 AGCCCCCTGCTGCCCGCAGACATGTGGACATGTGGACACCAAGTTCAGAGATACCTTGCAG 600
Db 909 AGCCCCCTGCTGCCCGCAGACATGTGGACATGTGGACACCAAGTTCAGAGATACCTTGCAG 968
QY 601 CGCCAGGCGCGTGGCGCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 660
Db 1029 CGCCAGGCGCGTGGCGCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1088
QY 721 TAGCGGCTGACTTTCGTCCTTGGGAGGAGCTTCAGGCGCTTCGAGATCAATTC 780
Db 1089 TAGCGGCTGACTTTCGTCCTTGGGAGGAGCTTCAGGCGCTTCGAGATCAATTC 1148
QY 781 AGCCCCCAACATGACCCGTCACCGCGTCCACCGCGTCCACCGCGTCCACCGCGTCCACCG 840
Db 1269 AGGCGAG 1274
```

RESULT 2

```
US-10-635-977-1
; Sequence 1, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635, 977
```

```
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-635-977-1

Query Match 100.0%; Score 906; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 1.8e-260;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGAC 60
Db 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGCGCGAGTGGGAGGAC 428
QY 61 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAAT 120
Db 429 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAAT 488
QY 121 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCACGTCGTGTGAACCTTCAGACGGAC 180
Db 489 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCACGTCGTGTGAACCTTCAGACGGAC 548
QY 181 ATTGACGGCTCCGGAAACATCTGGATATTAAGCCCGCCGCAAGTCCCGGGGCCGAGAC 240
Db 549 ATTGACGGCTCCGGAAACATCTGGATATTAAGCCCGCCGCAAGTCCCGGGGCCGAGAC 608
QY 241 ATAGTGTGATGACACGTCGTGGAGGAGATCCTGGAGCTGGCAGCTCGACACCAACCTCTT 300
Db 609 ATAGTGTGATGACACGTCGTGGAGGAGATCCTGGAGCTGGCAGCTCGACACCAACCTCTT 668
QY 301 TCCAGGGAACAAGTGGGTGCTCCAGAAATGATACATCGAGACGCCGCTGCTCATCTGTGAC 360
Db 669 TCCAGGGAACAAGTGGGTGCTCCAGAAATGATACATCGAGACGCCGCTGCTCATCTGTGAC 728
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGGACACCCCTGACCATCTGG 420
Db 729 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGGACACCCCTGACCATCTGG 788
QY 421 TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 480
Db 789 TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 848
QY 481 AGCGCCATCCACTGTGCAACAAACGCGCTCCAGAAATGATCTGAAGAAATGATGTGGGCCG 540
Db 849 AGCGCCATCCACTGTGCAACAAACGCGCTCCAGAAATGATCTGAAGAAATGATGTGGGCCG 908
QY 541 AGCCCCCTGCTGCCCGCAGACATGTGGACATGTGGACACCAAGTTCAGAGATACCTTGCAG 600
Db 909 AGCCCCCTGCTGCCCGCAGACATGTGGACATGTGGACACCAAGTTCAGAGATACCTTGCAG 968
QY 601 CGCCAGGCGCGTGGCGCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 660
Db 969 CGCCAGGCGCGTGGCGCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1028
QY 661 GCCACGCCATGAAGTGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 720
Db 720 GCCACGCCATGAAGTGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 788
QY 721 TAGCGGCTGACTTTCGTCCTTGGGAGGAGCTTCAGGCGCTTCGAGATCAATTC 780
Db 1089 TAGCGGCTGACTTTCGTCCTTGGGAGGAGCTTCAGGCGCTTCGAGATCAATTC 1148
QY 781 AGCCCCCAACATGACCCGTCACCGCGTCCACCGCGTCCACCGCGTCCACCGCGTCCACCG 840
Db 1269 AGGCGAG 1274
```



```
Db 1149 AGCCCCACCATGACCCGTCACCGCCGTCACGCCCGAGCTGTGTGCACAGGTGCAGGAG 1208
Qy 841 GACACCATCAAGGTGGCGGTGGACCGCAGCTGTGCATTCGGCAACTTCGAGCTCCTGTGG 900
Db 1209 GACACCATCAAGGTGGCGGTGGACCGCAGCTGTGCATTCGGCAACTTCGAGCTCCTGTGG 1268
Qy 901 AGGCAG 906
Db 1269 AGGCAG 1274

RESULT 3
US-10-615-659-12
; Sequence 12, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-12

Query Match 100.0%; Score 906; DB 19; Length 3554;
Best Local Similarity 100.0%; Pred. No. 2.1e-260;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGGACATCGACACGTGAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGTGGGAGGAC 60
Db 1911 GAGGACATCGACACGTGAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGTGGGAGGAC 1970
Qy 61 CTGACCCAGCAGTACTACTCTCCGTTCATGGCGATGTTTCATCTCCAATTCAGAAAT 120
Db 1971 CTGACCCAGCAGTACTACTCTCCGTTCATGGCGATGTTTCATCTCCAATTCAGAAAT 2030
Qy 121 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATACAGTCTGTGAACCTTCAGACGGAC 180
Db 2031 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATACAGTCTGTGAACCTTCAGACGGAC 2090
Qy 181 ATTGACGGGCTCCGGAACATCTGGAATATATAAGCCCGGCAAGTCCCGGGGCGGAGAC 240
Db 2091 ATTGACGGGCTCCGGAACATCTGGAATATATAAGCCCGGCAAGTCCCGGGGCGGAGAC 2150
Qy 241 ATAGTGTGCATGAGCAGTGTGGAGGATCCTTGAGCTGTCAGCTGTCAGACCACTCTT 300
Db 2151 ATAGTGTGCATGAGCAGTGTGGAGGATCCTTGAGCTGTCAGCTGTCAGACCACTCTT 2210
Qy 301 TCCAGGACCAACAGTGGGTGTCAGAAATGATACAGAGCGCGCTGCTCATCTGTGAC 360
Db 2211 TCCAGGACCAACAGTGGGTGTCAGAAATGATACAGAGCGCGCTGCTCATCTGTGAC 2270
Qy 361 ACCAAGTTCGACATCAGACAGTGTTCCTGTGTACAGGACTGGAACCCCTTCAGCATTGG 420
Db 2271 ACCAAGTTCGACATCAGACAGTGTTCCTGTGTACAGGACTGGAACCCCTTCAGCATTGG 2330
Qy 421 TTCTACAGGAGAGTACTTTCGGGTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 480
Db 2331 TTCTACAGGAGAGTACTTTCGGGTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 2390
Qy 481 AGGCCATCCACCTGTGTGCAACAGCCGTCAGAAATGATGAGAAATGATGTGGGCGC 540
Db 2391 AGGCCATCCACCTGTGTGCAACAGCCGTCAGAAATGATGAGAAATGATGTGGGCGC 2450
```

```
Qy 541 AGCCCCCTGCTGCCGCACACATGTGGACCAACAGGTTCACAGGAGTACCTGCAG 600
Db 2451 AGCCCCCTGCTGCCGCACACATGTGGACCAACAGGTTCACAGGAGTACCTGCAG 2510
Qy 601 CGCCAGGGCGGTGGCGCGGTGTGGGCGAGCGTCACTTACCCGTCATGAAGAAGGCCATC 660
Db 2511 CGCCAGGGCGGTGGCGCGGTGTGGGCGAGCGTCACTTACCCGTCATGAAGAAGGCCATC 2570
Qy 661 GCCCAGCCCATGAAGTGGCCCGAGGACCAAGTGGAGGCTCGCAAGACAGCTTTGAGCTC 720
Db 2571 GCCCAGCCCATGAAGTGGCCCGAGGACCAAGTGGAGGCTCGCAAGACAGCTTTGAGCTC 2630
Qy 721 TAGCGGCGTGAATTCCTTGGGAGGAGCTTCAGGGCCCTGGCTGTGATCGAGATCAATTC 780
Db 2631 TAGCGGCGTGAATTCCTTGGGAGGAGCTTCAGGGCCCTGGCTGTGATCGAGATCAATTC 2690
Qy 781 AGCCCCACCATGACCCGTCACCGCGTCACCGCCAGCTGTGTGCACAGGTGCAGGAG 840
Db 2691 AGCCCCACCATGACCCGTCACCGCGTCACCGCCAGCTGTGTGCACAGGTGCAGGAG 2750
Qy 841 GACACCATCAAGTGGCGGTGGACCGGAGCTGTGCATTCGGCAACTTCGAGCTCCTGTGG 900
Db 2751 GACACCATCAAGTGGCGGTGGACCGGAGCTGTGCATTCGGCAACTTCGAGCTCCTGTGG 2810
Qy 901 AGGCAG 906
Db 2811 AGGCAG 2816

RESULT 4
US-10-635-977-12
; Sequence 12, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-12

Query Match 100.0%; Score 906; DB 19; Length 3554;
Best Local Similarity 100.0%; Pred. No. 2.1e-260;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGGACATCGACACGTGAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGTGGGAGGAC 60
Db 1911 GAGGACATCGACACGTGAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGTGGGAGGAC 1970
Qy 61 CTGACCCAGCAGTACTACTCCCTGCTTCATGGCGATGTTTCATCTCCAATTCAGAAAT 120
Db 1971 CTGACCCAGCAGTACTACTCCCTGCTTCATGGCGATGTTTCATCTCCAATTCAGAAAT 2030
Qy 121 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATACAGTCTGTGAACCTTCAGACGGAC 180
Db 2031 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATACAGTCTGTGAACCTTCAGACGGAC 2090
Qy 181 ATTGACGGGCTCCGGAACATCTGGAATATATAAGCCCGGCAAGTCCCGGGGCGGAGAC 240
Db 2091 ATTGACGGGCTCCGGAACATCTGGAATATATAAGCCCGGCAAGTCCCGGGGCGGAGAC 2150
Qy 241 ATAGTGTGCATGAGCAGTGTGGAGGATCCTTGAGCTGTCAGCTGTCAGACCACTCTT 300
```


Db 2151 ATAGTGTGATGACCGTGTGGAGGAGATCCTGGAGCTGGACGTGCAGACCAACCCCTCTT 2210
|
Qy 301 TCCAGGGACAAAGTGGGTGTCAGAGTACATCGAGACCGCGTGTCTCTCTGTGAC 360
|
Db 2211 TCCAGGGACAAAGTGGGTGTCAGAGTACATCGAGACCGCGTGTCTCTCTGTGAC 2270
|
Qy 361 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAACGAGTGGAAACCCCTGACCATCTGG 420
|
Db 2271 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAACGAGTGGAAACCCCTGACCATCTGG 2330
|
Qy 421 TTCTACAAGAGAGTGTACTTGGCGGTCTCAACTCAGCGGTCTTCCCTTGGACAAAGCTGGAC 480
|
Db 2331 TTCTACAAGAGAGTGTACTTGGCGGTCTCAACTCAGCGGTCTTCCCTTGGACAAAGCTGGAC 2390
|
Qy 481 AGCGCCATCCACCTGTGCAACAAACGCGTCCAGAGTACCTGAAGAAATGATGTGGGCCCG 540
|
Db 2391 AGCGCCATCCACCTGTGCAACAAACGCGTCCAGAGTACCTGAAGAAATGATGTGGGCCCG 2450
|
Qy 541 AGCCCCCTGTGTCGCCGACACAAACATGTGGACCAAGAGTTCAGAGGAGTACCTGCGAG 600
|
Db 2451 AGCCCCCTGTGTCGCCGACACAAACATGTGGACCAAGAGTTCAGAGGAGTACCTGCGAG 2510
|
Qy 601 CGCCAGGGCGTGGCGCGTGTGGGCGAGGTCATCTACCCGTCATGAAAGAGGCCATC 660
|
Db 2511 CGCCAGGGCGTGGCGCGTGTGGGCGAGGTCATCTACCCGTCATGAAAGAGGCCATC 2570
|
Qy 661 GCCCAGCCATGAAGTGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTGGAGTTC 720
|
Db 2571 GCCCAGCCATGAAGTGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTGGAGTTC 2630
|
Qy 721 TAGCGGGCTGACTTCGTCTCTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTC 780
|
Db 2631 TAGCGGGCTGACTTCGTCTCTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTC 2690
|
Qy 781 AGCCCCCACCATGACCCGTCACCGCGTCCAGCGGTCAGCGCCAGCTGTGTGCACAGGTGCGAGGAG 840
|
Db 2691 AGCCCCCACCATGACCCGTCACCGCGTCCAGCGGTCAGCGCCAGCTGTGTGCACAGGTGCGAGGAG 2750
|
Qy 841 GACACCATCAAGTGGCGCTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
|
Db 2751 GACACCATCAAGTGGCGCTGGAGCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 2810
|
Qy 901 AGGCAG 906
|
Db 2811 AGGCAG 2816
|

RESULT 5
US-10-615-659-10
; Sequence 10, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-10

Query Match 89.1%; Score 806.8; DB 19; Length 1859;
Best Local Similarity 95.0%; Pred. No. 8.7e-231;
Matches 861; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

Qy 1 GAGGCATCGACACCGTCAAGATGCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAC 60
|
Db 263 GAGGCATCGACACCGTCAAGATGCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAC 322
|
Qy 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAGAAAT 120
|
Db 323 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAGAAAT 351
|
Qy 121 TACTTTTCGAGTGGCAGGCTCTGCTGAATAGAAATCACTGCTGTGAACCTCAGACGGAC 180
|
Db 352 -----TGCAGGCTCTGCTGAATAGAAATCACTGCTGTGAACCTCAGACGGAC 399
|
Qy 181 ATTGACGGGCTCCGGAACATCTGGATTATAAGCCCGCGCAAGTCCCGGGGCCAGAC 240
|
Db 400 ATTGACGGGCTCCGGAACATCTGGATTATAAGCCCGCGCAAGTCCCGGGGCCAGAC 459
|
Qy 241 ATAGTGTGATGACCGTGTGGAGGAGATCCTGGAGCTGGAGCTGCAGACCAACCCCTCTT 300
|
Db 460 ATAGTGTGATGACCGTGTGGAGGAGATCCTGGAGCTGGAGCTGCAGACCAACCCCTCTT 519
|
Qy 301 TCCAGGGACAAAGTGGGTGGTCCAGAAATACATCGAGACCGCGCTGTCTCTCTGTGAC 360
|
Db 520 TCCAGGGACAAAGTGGGTGGTCCAGAAATACATCGAGACCGCGCTGTCTCTCTGTGAC 579
|
Qy 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGAGTGGAAACCCCTGACCATCTGG 420
|
Db 580 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGAGTGGAAACCCCTGACCATCTGG 639
|
Qy 421 TTCTACAAGAGAGTACTTGGCGGTCTCAACTCAGCGCTTCTCCCTGACAGCTGGAC 480
|
Db 640 TTCTACAAGAGAGTACTTGGCGGTCTCAACTCAGCGCTTCTCCCTGACAGCTGGAC 699
|
Qy 481 AGCGCCATCCACTGTGCAACAAACGCGTCCAGAAATACATCGAGAAATGATGTGGGCCCG 540
|
Db 700 AGCGCCATCCACTGTGCAACAAACGCGTCCAGAAATGATGTGGGCCCG 759
|
Qy 541 AGCCCCCTGCTGCCGACACAAACATGTGGACCAAGTTCAGAGTACCTGCGAG 600
|
Db 760 AGCCCCCTGCTGCCGACACAAACATGTGGACCAAGTTCAGAGTACCTGCGAG 819
|
Qy 601 CGCCAGGGCGTGGCGCGTGTGGGCGAGCTCATCTACCCGTCATGAAAGAGGCCATC 660
|
Db 820 CGCCAGGGCGTGGCGCGTGTGGGCGAGCTCATCTACCCGTCATGAAAGAGGCCATC 879
|
Qy 661 GCCCAGCCATGAAGTGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTGGAGTTC 720
|
Db 880 GCCCAGCCATGAAGTGGCCAGGACCAAGTGGAGCCCTCGCAAGAACAGCTTGGAGTTC 939
|
Qy 721 TAGCGGGCTGACTTCGTCTCTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTC 780
|
Db 940 TAGCGGGCTGACTTCGTCTCTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTC 999
|
Qy 781 AGCCCCCACCATGACCCGTCACCGCGTCCAGGCGGCTGAGTGCACAGGTGCGAGGAG 840
|
Db 1000 AGCCCCCACCATGACCCGTCACCGCGGTCAGGCGGCTGAGTGCACAGGTGCGAGGAG 1059
|
Qy 841 GACACCATCAAGTGGCGCTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
|
Db 1060 GACACCATCAAGTGGCGCTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 1119
|
Qy 901 AGGCAG 906
|
Db 1120 AGCCCG 1125
|

RESULT 6
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42


```

; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-635-977-10

Query Match      89.1%; Score 806.8; DB 19; Length 1859;
Best Local Similarity 95.0%; Pred. No. 8.7e-231;
Matches 861; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

QY 1 GAGGACATCGACAGCTCAGCAGATGCCGTGGAGGAGCTCACTGAGCGCGAGTGGGAGGAC 60
DB 263 GAGGACATCGACAGCTCAGCAGATGCCGTGGAGGAGCTCACTGAGCGCGAGTGGGAGGAC 322
QY 61 CTGACCCAGCAGTACTACCTCCCTCGTTCATGGCGATGCTTTTCATCTCCAATTCAGAAAT 120
DB 323 CTGACCCAGCAGTACTACCTCCCTCGTTCATGGCGATGCTTTTCATCTCCAATTCAGAAAT 351
QY 121 TACTTTTCCGAGTCCAGGCTCTGCTGAATAGAAATCACTGCTGTGTGAACCTTCAGACGGAC 180
DB 352 -----TGCCAGGCTCTGCTGAATAGAAATCACTGCTGTGTGAACCTTCAGACGGAC 399
QY 181 ATTGACGGGCTCCGGAAACATCTCGAATATTAAGCCCGCGGCGCAAGTCCCGGGGCGGAGAC 240
DB 400 ATTGACGGGCTCCGGAAACATCTCGAATATTAAGCCCGCGGCGCAAGTCCCGGGGCGGAGAC 459
QY 241 ATAGTGTGCATGACACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCACTCTT 300
DB 460 ATAGTGTGCATGACACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCACTCTT 519
QY 301 TCCAGGAGCAACAAGTGGGTGGTCCAGAAAGTATCATCGAGACGCGCTGCTCATCTGTGAC 360
DB 520 TCCAGGAGCAACAAGTGGGTGGTCCAGAAAGTATCATCGAGACGCGCTGCTCATCTGTGAC 579
QY 361 ACCAAGTTGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTGACCAATCTGG 420
DB 580 ACCAAGTTGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTGACCAATCTGG 639
QY 421 TTCTACAGGAGATTTACTTGGGTTCTCACTCAGCGCTTCTCCCTGGACAAAGCTGGAC 480
DB 640 TTCTACAGGAGATTTACTTGGGTTCTCACTCAGCGCTTCTCCCTGGACAAAGCTGGAC 699
QY 481 AGCGCCATCCACCTGTGCAACACGCGTCCAGAAATGCTGAAAGTATGTTGGGCGCG 540
DB 700 AGCGCCATCCACCTGTGCAACACGCGTCCAGAAATGCTGAAAGTATGTTGGGCGCG 759
QY 541 AGCCCCCTGTGCGCGGACACAACAATGTGGACACGACGAGTTCACGAGTACCTGACG 600
DB 760 AGCCCCCTGTGCGCGGACACAACAATGTGGACACGACGAGTTCACGAGTACCTGACG 819
QY 601 CGCAGGCGCGTGGCGCGTGTGGGCGAGGCTCATCTACCGTCCATGAAAGAGGCGCATC 660
DB 820 CGCAGGCGCGTGGCGCGTGTGGGCGAGGCTCATCTACCGTCCATGAAAGAGGCGCATC 879
QY 661 GCCACGCCATGAAGTGGCGGAGGACGCTGAGGAGCTCGCAAGAAAGCTTTGAGCTC 720
DB 880 GCCACGCCATGAAGTGGCGGAGGACGCTGAGGAGCTCGCAAGAAAGCTTTGAGCTC 939
QY 721 TACGGGCTGACCTTCGCTCTGGGAGGAGCTTCAGGCGCTTGGCTGATCGAGATCAATCC 780
DB 940 TACGGGCTGACCTTCGCTCTGGGAGGAGCTTCAGGCGCTTGGCTGATCGAGATCAATCC 999
QY 781 AGCCCCACCATGACCCGCTCCAGCGGCTCA CGGCGCTGACAGCTGTGTGCA CAGGTGCGAGG 840

```


2367	Db		AGCCCCGTGTGCCCGCACAAATGTGGACCAAGCCAGGTTCCAGGAGTACCTCGAG	2426
601	Qy		CGCAGAGGCGGTGCGCCGTGTGGGCGAGCGTCACTACCGTCCATGAAGAGGCCATC	660
2427	Db		CGCCAGGCGGTGCGCCGTGTGGGCGAGCGTCACTACCGTCCATGAAGAGGCCATC	2486
661	Qy		GCCACGCCATGAAGTGGCCAGGACCACTGGAGCCTCGCAAGAACACGCTTTGAGCTC	720
2487	Db		GCCACGCCATGAAGTGGCCAGGACCACTGGAGCCTCGCAAGAACACGCTTTGAGCTC	2546
721	Qy		TACGGGCTGACTTCGTCCTTTGGAGGGACTTACAGGCCCTGGCTGATCGAGATCAATTC	780
2547	Db		TACGGGCTGACTTCGTCCTTTGGAGGGACTTACAGGCCCTGGCTGATCGAGATCAATTC	2606
781	Qy		AGCCCCACCATGCAACCGGTCCAACGCCCGTCAACGCCCGTGTGTGTCACAGGTGCAGGAG	840
2607	Db		AGCCCCACCATGCAACCGGTCCAACGCCCGTCAACGCCCGTGTGTGTCACAGGTGCAGGAG	2666
841	Qy		GACACCATCAAGGTGGCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG	900
2667	Db		GACACCATCAAGGTGGCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG	2726
901	Qy		AGGCAG 906	
2727	Db		AGGCAG 2732	

```

RESULT 8
US-10-635-977-11
; Sequence 11, Application US/10635977
; Publication No. US2004071131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-11

```

Db	2130	-----2126	
Qy	301	TCAGGGAACAAGTGGTGGTCCAGAAGTACATCGAGACGCGCTGCTCATCTGTGAC	360
Db	2130	---AGGGAACAAGTGGTGGTCCAGAAGTACATCGAGACGCGCTGCTCATCTGTGAC	2186
Qy	361	ACCAAGTTCACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTGACCAATCGG	420
Db	2187	ACCAAGTTCACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTGACCAATCGG	2246
Qy	421	TTCTACAAGGAGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC	480
Db	2247	TTCTACAAGGAGAGTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC	2306
Qy	481	AGGCCATCACCTGTGCAACAACGCGTCCAGAAGTACTGAAAGAAATGATGTGGGCGC	540
Db	2307	AGGCCATCACCTGTGCAACAACGCGTCCAGAAGTACTGAAAGAAATGATGTGGGCGC	2366
Qy	541	AGCCCCGTGTCGCCGACACACATGTGACACAGACACAGGTTCACAGGAGTACCTGCAG	600
Db	2367	AGCCCCGTGTCGCCGACACACATGTGACACAGACACAGGTTCACAGGAGTACCTGCAG	2426
Qy	601	CGCCAGGCGCTGGCGCGTGTGGGCGAGGCTCATCTACCCGTCCTATGAAGAAGGCCATC	660
Db	2427	CGCCAGGCGCTGGCGCGTGTGGGCGAGGCTCATCTACCCGTCCTATGAAGAAGGCCATC	2486
Qy	661	GCCACGCCATGAAGTGGGCCAGGACCAAGTGGAGGCTCGCAAGAACAGCTTTGAGCTC	720
Db	2487	GCCACGCCATGAAGTGGGCCAGGACCAAGTGGAGGCTCGCAAGAACAGCTTTGAGCTC	2546
Qy	721	TACGGGCTGACTTCGTTCCTTGGGAGGGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC	780
Db	2547	TACGGGCTGACTTCGTTCCTTGGGAGGGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC	2606
Qy	781	AGCCCCACCATGACCCGTCACGCGGTCACGCGCCAGCTGTGTGCA CAGGTGCAGGAG	840
Db	2607	AGCCCCACCATGACCCGTCACGCGGTCACGCGCCAGCTGTGTGCA CAGGTGCAGGAG	2666
Qy	841	GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG	900
Db	2667	GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG	2726
Qy	901	AGGCAG 906	
Db	2727	AGGCAG 2732	

```

RESULT 9
US-10-615-659-9
; Sequence 9, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-9

```

Query Match 83.8%; Score 760.4; DB 19; Length 1939;
Best Local Similarity 92.5%; Pred. No. 6.8e-217;
Matches 838; Conservative 0; Mismatches 1; Indels 67; Gaps 1;
QV 1 GAGGACATCAGACGTCAGCAGATGCGTCGGAGGACCTCAGTCAGGCCGAGTGGGAGGAC 60

Db 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGACCTCACTGAGGCGGAGTGGAGGAC 428
Qy 61 CTGACCCAGCAGTACTACTCTCGTTCATGCGGATGCTTTCATCTCAATTCAGAAAT 120
Db 429 CTGACCCAGCAGTACTACTCTCGTTCATGCGGATGCTTTCATCTCAATTCAGAAAT 488
Qy 121 TACTTTTCGACGTCCAGGCTCTCTGAATAGATCAGCTCTGTGAACCTTCACACGGAC 180
Db 489 TACTTTTCGACGTCCAGGCTCTCTGAATAGATCAGCTCTGTGAACCTTCACACGGAC 548
Qy 181 ATTGACGGGCTCCGGAACATCTGATATATAAGCCCGCGGCAAGTCCCGGGGCGG--- 604
Db 549 ATTGACGGGCTCCGGAACATCTGATATATAAGCCCGCGGCAAGTCCCGGGGCGG--- 604
Qy 241 ATAGTGTGCATGACCGTGTGGAGAGATCTCTGAGCTGGCAGCTGCGAGCACCCCTCTT 300
Db 605 ----- 604
Qy 301 TCCAGGACACAAAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCTCATCTGTGAC 360
Db 605 ---AGGACACAAAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCTCATCTGTGAC 661
Qy 361 ACCAAGTTCGACATCAGACAGTGGTTCCTGTCAAGGACTGGAAACCCCTGACCATCTGG 420
Db 662 ACCAAGTTCGACATCAGACAGTGGTTCCTGTCAAGGACTGGAAACCCCTGACCATCTGG 721
Qy 421 TTCTACAAGGAGAGTTACTTTCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGAC 480
Db 722 TTCTACAAGGAGAGTTACTTTCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGAC 781
Qy 481 AGCGCCATCCACTGTGTCGCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC 780
Db 1022 TAGCGGCTGACTTCGCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC 1081
Qy 781 AGCCCCACCATGACCCGTCACCGCGGTCAAGGCGGCTGAGTGTGACAGGTCAGGAG 840
Db 1082 AGCCCCACCATGACCCGTCACCGCGGTCAAGGCGGCTGAGTGTGACAGGTCAGGAG 1141
Qy 841 GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 900
Db 1142 GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 1201
Qy 901 AGGCAG 906
Db 1202 AGGCAG 1207

RESULT 10
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BG542
; FILE REFERENCE: D0283A CIP

; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match 83.9%; Score 760.4; DB 19; Length 1939;
Best Local Similarity 92.5%; Pred. No. 6.8e-217;
Matches 838; Conservative 0; Mismatches 1; Indels 67; Gaps 1;
Qy 1 GAGGACATCGACACGTCAGCAGATGCCGTGGAGACCTCACTGAGGCGGAGTGGAGGAC 60
Db 369 GAGGACATCGACACGTCAGCAGATGCCGTGGAGACCTCACTGAGGCGGAGTGGAGGAC 428
Qy 61 CTGACCCAGCAGTACTACTCTCGTTCATGCGGATGCTTTCATCTCAATTCAGAAAT 120
Db 429 CTGACCCAGCAGTACTACTCTCGTTCATGCGGATGCTTTCATCTCAATTCAGAAAT 488
Qy 121 TACTTTTCGACGTCCAGGCTCTCTGAATAGATCAGCTCTGTGAACCTTCACACGGAC 180
Db 489 TACTTTTCGACGTCCAGGCTCTCTGAATAGATCAGCTCTGTGAACCTTCACACGGAC 548
Qy 181 ATTGACGGGCTCCGGAACATCTGATATATAAGCCCGCGGCAAGTCCCGGGGCGG--- 604
Db 549 ATTGACGGGCTCCGGAACATCTGATATATAAGCCCGCGGCAAGTCCCGGGGCGG--- 604
Qy 241 ATAGTGTGCATGACCGTGTGGAGAGATCTCTGAGCTGGCAGCTGCGAGCACCCCTCTT 300
Db 605 ----- 604
Qy 301 TCCAGGACACAAAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCTCATCTGTGAC 360
Db 605 ---AGGACACAAAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCTCATCTGTGAC 661
Qy 361 ACCAAGTTCGACATCAGACAGTGGTTCCTGTCAAGGACTGGAAACCCCTGACCATCTGG 420
Db 662 ACCAAGTTCGACATCAGACAGTGGTTCCTGTCAAGGACTGGAAACCCCTGACCATCTGG 721
Qy 421 TTCTACAAGGAGAGTTACTTTCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGAC 480
Db 722 TTCTACAAGGAGAGTTACTTTCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGAC 781
Qy 481 AGCGCCATCCACTGTGTCGCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC 780
Db 782 AGCGCCATCCACTGTGTCGCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC 841
Qy 541 AGCGCCATCCACTGTGTCGCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC 600
Db 842 AGCGCCATCCACTGTGTCGCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC 901
Qy 601 CGCAGGCGCGTGGCGCGGTGTGGGCGAGGCTCATCTACCGCTCCATGAAAGAGGCCATC 660
Db 902 CGCAGGCGCGTGGCGCGGTGTGGGCGAGGCTCATCTACCGCTCCATGAAAGAGGCCATC 961
Qy 661 GCCCAACCCATGAAGTGGCCAGGACCAAGTGGAGCTTCCAGGAGTACCTGAGTCC 720
Db 962 GCCCAACCCATGAAGTGGCCAGGACCAAGTGGAGCTTCCAGGAGTACCTGAGTCC 1021
Qy 721 TAGCGGCTGACTTCGCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC 780
Db 1022 TAGCGGCTGACTTCGCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTCC 1081
Qy 781 AGCCCCACCATGACCCGTCACCGCGGTCAAGGCGGCTGAGTGTGACAGGTCAGGAG 840
Db 1082 AGCCCCACCATGACCCGTCACCGCGGTCAAGGCGGCTGAGTGTGACAGGTCAGGAG 1141
Qy 901 AGGCAG 906
Db 1202 AGGCAG 1207


```
QY      841  GACACCATCAAGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
      |||||||
Db      1142  GACACCATCAAGTGGCCCGTGGACCGCAGCTGTGATCGGCAACTTCGAGCTCCTGTGG 1201
      |||||||
QY      901  AGGCAG 906
      |||||||
Db      1202  AGGCAG 1207
      |||||||

RESULT 11
US-10-615-659-3
; Sequence 3, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-615-659-3

Query Match      80.1%; Score 726; DB 19; Length 726;
Best Local Similarity 100.0%; Pred. No. 1e-206;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      181  ATTGACGGGCTCCGGAAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 240
      |||||||
Db      1      ATTGACGGGCTCCGGAAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 60
      |||||||
QY      241  ATAGTGTGATGACCCGCTGTGGAGGAGATCTTGAGCTGGCAGCTGCAGACCACTCTCTT 300
      |||||||
Db      61  ATAGTGTGATGACCCGCTGTGGAGGAGATCTTGAGCTGGCAGCTGCAGACCACTCTCTT 120
      |||||||
QY      301  TCCAGGGACAAACAGTGGGTGTCAGAAAGTACATCGAGACGCGCTGCTCATCTGTGAC 360
      |||||||
Db      121  TCCAGGGACAAACAGTGGGTGTCAGAAAGTACATCGAGACGCGCTGCTCATCTGTGAC 180
      |||||||
QY      361  ACCAAGTTCGACATCAGACAGTGGTTCTCCTCGTCA CGGACTGGAACCCCTGACCATCTGG 420
      |||||||
Db      181  ACCAAGTTCGACATCAGACAGTGGTTCTCCTCGTCA CGGACTGGAACCCCTGACCATCTGG 240
      |||||||
QY      421  TTCTAAGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 480
      |||||||
Db      481  AGCGCCATCCACTGTGCAACAAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGCCCGC 540
      |||||||
QY      541  AGCGCCATCCACTGTGCAACAAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGCCCGC 600
      |||||||
Db      361  AGCGCCATCCACTGTGCAACAAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGCCCGC 420
      |||||||
QY      601  CGCAGGGCGGTGGCCCGTGTGGGCGAGCGTCACTACCGGTCATGAAGAAAGGCCATC 660
      |||||||
Db      421  CGCAGGGCGGTGGCCCGTGTGGGCGAGCGTCACTACCGGTCATGAAGAAAGGCCATC 480
      |||||||
QY      661  GCCCAGCCATGAAGTGGCCCGAGGACCGCTGCGAGGACCTCGCAAGAACAGCTTGA3CTC 720
      |||||||
Db      481  GCCCAGCCATGAAGTGGCCCGAGGACCGCTGCGAGGACCTCGCAAGAACAGCTTGA3CTC 540
      |||||||
```

```
QY      721  TAGGGGGCTGACTTCGTCTCTTGGAGGGACTTTCAGGCCCTGGCTGATCGAGATCAATTCC 780
      |||||||
Db      541  TAGGGGGCTGACTTCGTCTCTTGGAGGGACTTTCAGGCCCTGGCTGATCGAGATCAATTCC 600
      |||||||
QY      781  AGCCCCACCATGACACCCGTCACGCCGCTCAGGCCCCAGCTGTGTGCACAGGTGCAGGAG 840
      |||||||
Db      601  AGCCCCACCATGACACCCGTCACGCCGCTCAGGCCCCAGCTGTGTGCACAGGTGCAGGAG 660
      |||||||
QY      841  GACACCATCAAGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
      |||||||
Db      661  GACACCATCAAGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 720
      |||||||
QY      901  AGGCAG 906
      |||||||
Db      721  AGGCAG 726
      |||||||

RESULT 12
US-10-635-977-3
; Sequence 3, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-635-977-3

Query Match      80.1%; Score 726; DB 19; Length 726;
Best Local Similarity 100.0%; Pred. No. 1e-206;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      181  ATTGACGGGCTCCGGAAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 240
      |||||||
Db      1      ATTGACGGGCTCCGGAAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 60
      |||||||
QY      241  ATAGTGTGATGACCCGCTGTGGAGGAGATCTTGAGCTGGCAGCTGCAGACCACTCTCTT 300
      |||||||
Db      61  ATAGTGTGATGACCCGCTGTGGAGGAGATCTTGAGCTGGCAGCTGCAGACCACTCTCTT 120
      |||||||
QY      301  TCCAGGGACAAACAGTGGGTGTCAGAAAGTACATCGAGACGCGCTGCTCATCTGTGAC 360
      |||||||
Db      121  TCCAGGGACAAACAGTGGGTGTCAGAAAGTACATCGAGACGCGCTGCTCATCTGTGAC 180
      |||||||
QY      361  ACCAAGTTCGACATCAGACAGTGGTTCTCCTCGTCA CGGACTGGAACCCCTGACCATCTGG 420
      |||||||
Db      181  ACCAAGTTCGACATCAGACAGTGGTTCTCCTCGTCA CGGACTGGAACCCCTGACCATCTGG 240
      |||||||
QY      421  TTCTAAGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 480
      |||||||
Db      241  TTCTAAGAGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 300
      |||||||
QY      481  AGCGCCATCCACTGTGCAACAAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGCCCGC 540
      |||||||
Db      301  AGCGCCATCCACTGTGCAACAAACCGCGTCCAGAGTACCTGAAGAAATGATGTGGCCCGC 360
      |||||||
QY      541  AGCGCCCTCTGCTCCCGCACACAAACATGTGGAACGACCAAGGTTCCAGGAGTACCTGCGAG 600
      |||||||
```


Db 361 AGCCCCCTGCTGCCCGCACACACATGTGACACGACACGAGTTCCAGGAGTACCTGCAG 420
Qy 601 CGCCAGGGCGGTGGCGCGGTGTGGGAGCGGTCACTACCGTCCATGAAGAAGGCCATC 660
Db 421 CGCCAGGGCGGTGGCGCGGTGTGGGAGCGGTCACTACCGTCCATGAAGAAGGCCATC 480
Qy 661 GCCACGCCATGAGGTGGGCCAGGACGAGTGGAGCCTGCAGACAGCTTTGAGCTC 720
Db 481 GCCACGCCATGAGGTGGGCCAGGACGAGTGGAGCCTGCAGACAGCTTTGAGCTC 540
Qy 721 TAGGGGCTGACTTCGTTCCTTGGGAGGAGCTTCAAGGCCCTGGGTGATCGAGATCAATTC 780
Db 541 TAGGGGCTGACTTCGTTCCTTGGGAGGAGCTTCAAGGCCCTGGGTGATCGAGATCAATTC 600
Qy 781 AGCCCCACCATGACCCGTCACGCCCGTCAACGCCCGAGCTGTGTGCACAGGTGCAAGGAG 840
Db 601 AGCCCCACCATGACCCGTCACGCCCGTCAACGCCCGAGCTGTGTGCACAGGTGCAAGGAG 660
Qy 841 GACACCATGAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 900
Db 661 GACACCATGAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG 720
Qy 901 AGGCAG 906
Db 721 AGGCAG 726
RESULT 13
US-10-029-386-24894/c
; Sequence 24894, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24894
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022327.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.35
; OTHER INFORMATION: SWISSPROT HIT: P38160, EVALUE 5.00e-04
; OTHER INFORMATION: EST HUMAN HIT: BE82753.1, EVALUE 3.00e-06
; OTHER INFORMATION: NT HIT: g114726145, EVALUE 8.00e-09
US-10-029-386-24894
Query Match 47.1%; Score 427; DB 16; Length 490;
Best Local Similarity 100.0%; Pred. No. 2.8e-117;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 480 CAGCGCCATCCACTGTGCAACACGCGTCCAGAGTACCTGAAGAATGATGTGGCGG 539
Db 427 CAGCGCCATCCACTGTGCAACACGCGTCCAGAGTACCTGAAGAATGATGTGGCGG 368
Qy 540 CAGCCCCCTGCTGCCCGCACACAACTGTGGACAGCACCAGGTTCAGGAGTACCTGCA 599
Db 367 CAGCCCCCTGCTGCCCGCACACAACTGTGGACAGCACCAGGTTCAGGAGTACCTGCA 308
Qy 600 CGCCAGGGCGGTGGCGCGGTGTGGGAGCGGTCACTACCGTCCATGAAGAAGGCCAT 659
Db 307 CGCCAGGGCGGTGGCGCGGTGTGGGAGCGGTCACTACCGTCCATGAAGAAGGCCAT 248
Qy 660 CGCCACGCCATGAGTGGCGCGGACGACGCTGAGCTTCGCAAGACAGCTTTGAGCT 719

Db 247 CGCCACGCCATGAGGTGGCGCGCAGGACCAACGCTGGAGCCTCGCAAGACAGCTTTGAGCT 188
Qy 720 CTACGGGGGTGACTTCGTTCCTTGGGAGGAGCTTTCAGGCCCTTGCTGATCGAGATCAATTC 779
Db 187 CTACGGGGGTGACTTCGTTCCTTGGGAGGAGCTTTCAGGCCCTTGCTGATCGAGATCAATTC 128
Qy 780 CAGCCCCACCATGACCCGTCACCGCGGTTCACGGGCCAGCTGTGTGCACAGGTGCAAGGA 839
Db 127 CAGCCCCACCATGACCCGTCACCGCGGTTCACGGGCCAGCTGTGTGCACAGGTGCAAGGA 68
Qy 840 GGACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCAGAGCTCCTGTG 899
Db 67 GGACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCAGAGCTCCTGTG 8
Qy 900 GAGGCAG 906
Db 7 GAGGCAG 1
RESULT 14
US-10-723-860-631/c
; Sequence 631, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 631
; LENGTH: 101270
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-631
Query Match 47.1%; Score 427; DB 20; Length 101270;
Best Local Similarity 100.0%; Pred. No. 1e-116;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 480 CAGCGCCATCCACTGTGCAACAAACGCGTCCAGAGTACCTGAAGAATGATGTGGCGG 539
Db 44821 CAGCGCCATCCACTGTGCAACAAACGCGTCCAGAGTACCTGAAGAATGATGTGGCGG 44762
Qy 540 CAGCCCCCTGCTGCCCGCACACAACTGTGGACCAAGGTTCAGGAGTACCTGCA 599
Db 44761 CAGCCCCCTGCTGCCCGCACACAACTGTGGACCAAGGTTCAGGAGTACCTGCA 44702
Qy 600 CGCCAGGGCGGTGGCGCGGTGTGGGAGCGCTCATCTACCGTCCATGAGAGGCCAT 659
Db 44701 CGCCAGGGCGGTGGCGCGGTGTGGGAGCGCTCATCTACCGTCCATGAGAGGCCAT 44642
Qy 660 CGCCACGCCATGAGGTGGCGCGCAGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCT 719
Db 44641 CGCCACGCCATGAGGTGGCGCGCAGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCT 44582
Qy 720 CTACGGGGGTGACTTCGTTCCTTGGGAGGAGCTTTCAGGCCCTTGCTGATCGAGATCAATTC 779
Db 44581 CTACGGGGGTGACTTCGTTCCTTGGGAGGAGCTTTCAGGCCCTTGCTGATCGAGATCAATTC 44522
Qy 780 CAGCCCCACCATGACCCCGTCACCGCGGTTCACGGGCCAGCTGTGTGCACAGGTGCAAGGA 839
Db 44521 CAGCCCCACCATGACCCCGTCACCGCGGTTCACGGGCCAGCTGTGTGCACAGGTGCAAGGA 44462
Qy 840 GGACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTG 899
Db 44461 GGACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGTG 44402

QY 900 GAGGACAG 906
Db 44401 GAGGACAG 44395

RESULT 15

US-10-275-595A-39
; Sequence 39, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, DYUNG Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BATRA, Sajeev
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 2380
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2156553CB1
US-10-275-595A-39

Query Match 44.0%; Score 398.8; DB 18; Length 2380;
Best Local Similarity 66.3%; Pred. No. 1.1e-108;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY 3 GGACATCGACAGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGGAGACCT 62
Db 473 GGACATCGACAGGACCTGGAGGCCCGCTGTACTCTCACCCCGAGGGCTGCTCCCTCTT 532
QY 63 GACCAGCAGTACTACTCCTCGTTTCAGCGATGCTTTCTCATCTCCAATTCAAGAAATTA 122
Db 533 CCTCAGCGCCTACTACCAAGTGGTCCACGAAGGGCAGAACTCAGGCACCTCGACACTCA 592
QY 123 CTTTTCGCAAGTCGAGGCTCTGTGAATAGAAATCAAGTCTGTGAACCTTCAGACGGACAT 182
Db 593 GGTCCAGCGCTGTGAGGACATCTTCAGCAGCTGCAGGCCGTGGTATCCCCAGATAGACAT 652
QY 183 TGACGGGCTCCGGAACATCTGGATTATAAGCCGGGGCCAAAGTCCCGGGCCCGAGACAT 242
Db 653 GGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGTGGACGAGGCAT 712
QY 243 AGTGTGCATGGACCGGTGTCGAGAGATCTGTGGAGCTGGCAGCTGCAGACCCCTCTTTC 302
Db 713 CATGTGATGGACCACTCGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGTAT 772

QY 303 CAGGGACAACAAAGTGGGTGGTCCAGAAAGTACATCGAGAGCGCGCTGCTCATCTGTGACAC 362
Db 773 GAAGGACGGCAAGTGGGTGGTGCAGAAAGTATATTGAGCGGCCCTCTCATCTTTGGCAC 832
QY 363 CAAGTTGACATCAGACAGTGGTTCCTCGTCAAGGACTGGAAACCCCTCAACCATCTGGTT 422
Db 833 CAAAGTTTGACCTCAGACAGTGGTTCCTCGTAACTGACTGGAACCCCATTTACCGTGTGGTT 892
QY 423 CTACAAGGAGAGTAACTTTCGGGTTCCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAG 482
Db 893 CTACCGGACAGCTATATCCGGTTTTTCCACGAGGCCCTTCTCCCTGAAGAACTCGACAA 952
QY 483 CGCCATCCACCTGTGTCAACCAACCGCTCCAGAAAGTACCTGAAGAATGATGTGGGCCGAG 542
Db 953 CTCAGTGACCTGTGTCAACCAACTCCAGAAGCACTGGAGAACTCATGCGCATCGGCA 1012
QY 543 CCCCCTGTGCCCGCACACAACATGTGGACAGCAGGTTCCAGGAGTACTCTGCAGCG 602
Db 1013 TCCACTGTCTCCGCCAGACAACATGTGGTCTAGCCAGAGGTTCCAGGCCCACTTCGAGGA 1072
QY 603 CCAGGGCGGTGGCGCGGTGTGGGGCAGCGTCACTACCGTCCATGAAGAAGCCATCGC 662
Db 1073 GATGGGTGCCCCAAATGCTTGGTCCACATCATCGTGCCTGCGATGAAGGATGCTGTGAT 1132
QY 663 CCACGCCATGAAGGTGGCCAGGACCAAGTGGAGCCTCGCAAGAAACAGCTTTGAGCTCTA 722
Db 1133 CCACGCACTTTCAGACCTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTA 1192
QY 723 CGGGGCTGACTTCGTCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTCAG 782
Db 1193 TGGCGCTGACTTTCGTGTTCGGGGAGGACTTCCAGGCCCTGGCTGATTTGAGATCAAAGCCAG 1252
QY 783 CCCCAACCATGCAACCCGTCCAGCCCGTCAAGCCCGTGTGTGACACAGGTGCAGGAGGA 842
Db 1253 CCCCAAGATGGCACCTCCACAGCAGTCACTGCGCGGCTCTGTGTGGCGTGCAGACTGA 1312
QY 843 CACCATCAAGGTGGCGGTGGACCGCA 868
Db 1313 CACCCTGGCGGTGGTCAATTGACCGGA 1338

Search completed: September 24, 2005, 16:48:43
Job time : 701.935 secs

Result No.	Score	Query		DB	ID	Description	
		Match	Length				
1	398.8	44.0	3001	4	US-09-620-312D-494	Sequence 494, Appl	
2	52.6	5.8	1272	4	US-09-252-991A-2471	Sequence 2471, Ap	
C	3	52.6	5.8	1449	4	US-09-252-991A-3200	Sequence 2200, Ap
	4	52.6	5.8	2797	1	US-09-252-991A-2632	Sequence 2632, Ap
	5	48.6	5.4	2277	4	US-08-676-967-5	Sequence 5, Appl
6	48.6	5.4	2277	1	US-08-676-974-5	Sequence 5, Appl	
7	48.6	5.4	2277	2	US-09-098-487-5	Sequence 5, Appl	
8	48.2	5.3	2004	1	US-08-471-033-18	Sequence 18, Appl	
9	48.2	5.3	2004	2	US-08-471-044-18	Sequence 18, Appl	
10	48.2	5.3	2004	2	US-08-463-483A-18	Sequence 18, Appl	
11	48.2	5.3	2004	2	US-08-471-046A-18	Sequence 18, Appl	
12	48.2	5.3	2004	2	US-08-470-566B-18	Sequence 18, Appl	
13	48.2	5.3	2004	2	US-08-469-334-18	Sequence 18, Appl	
14	48.2	5.3	2004	3	US-09-300-529-18	Sequence 18, Appl	
15	48.2	5.3	2655	1	US-08-471-033-17	Sequence 17, Appl	
16	48.2	5.3	2655	2	US-08-471-044-17	Sequence 17, Appl	
17	48.2	5.3	2655	2	US-08-463-483A-17	Sequence 17, Appl	
18	48.2	5.3	2655	2	US-08-471-046A-17	Sequence 17, Appl	
19	48.2	5.3	2655	2	US-08-470-566B-17	Sequence 17, Appl	
20	48.2	5.3	2655	2	US-08-469-334-17	Sequence 17, Appl	
21	48.2	5.3	2655	3	US-09-300-529-17	Sequence 17, Appl	
22	48	5.3	5314	4	US-09-620-312D-138	Sequence 138, App	
23	47.6	5.3	4884	4	US-09-252-991A-12126	Sequence 12126, A	
C	24	47.6	5.3	4884	4	US-09-252-991A-121292	Sequence 121292, A
	25	45.8	5.1	1515	4	US-09-502-540-9184	Sequence 9184, Ap
C	26	45.8	5.1	9556	4	US-09-902-540-982	Sequence 982, App
	27	45.6	5.0	1047	4	US-09-252-991A-14678	Sequence 14678, A


```
QY 63 GACCCAGCAGTACTACTCCCTCGTTTCATGCGCATGCTTTTCATCTCCAAATCAAGAAATTA 122
Db      |||||
1423 CCTCCAGCGCTACTCAAGTGGTCCACGAAGGGCAGAACTCAGGCACCTCGACACTCA 1482
QY 123 CTTTTCGAGTCCAGGCTCTGCTGTAATAGATCACTGTCGTGAAACCTTCAGACGACAT 182
Db      |||||
1483 GGTCCAGCGCTGTGAGGACATCTTCGACGACGTGTCAGGCGGTGGTACCCCGAGATGACAT 1542
QY 183 TGACGGGCTCCGGAACATCTGGATTTATAAGCCCGCGGCAAGTCCCGGGCGCGAGACAT 242
Db      |||||
1543 GGAAGGGATCGCAACATCTGGATCTGTAAGCCAGGAGCCAGTCCCGCGGAGGAGAT 1602
QY 243 AGTGTGCATGGACCGTGTGAGGAGATCTCTGAGCTGGCAGCTGCAGACACCCCTCTTTC 302
Db      |||||
1603 CATGTGCATGGACCACTCGAGGAGATGCTGAAGCTGGTGAACCGCAACCCCGTGTGAT 1662
QY 303 CAGGGACACAGTGGTGGTCCAGAGTACATCTGAGACGCCGCTCTCATCTGTGTGACAC 362
Db      |||||
1663 GAAGGACCGCAGTGGTGGTGCAGAAATATATGAGCGGCGCCCTCTCATCTTTGGCAC 1722
QY 363 CAAATTCGACATCAGACAGTGGTTCCTCGTCAAGGACTGGAAACCCCTCGACATCTGGTT 422
Db      |||||
1723 CAAATTTGACCTCAGACAGTGGTTCCTGTTAACTGACTGGAACCACTTACCGTGTGGTT 1782
QY 423 CTACAAAGAGAGTACTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAG 482
Db      |||||
1783 CTACCGCAGACAGTATATCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCTGGACAA 1842
QY 483 CGCCATCCACCTGTGCAACAAACGCCGTCCAGAGTACCTGAAGAAATGATGTGGGCCGAG 542
Db      |||||
1843 CTCAGTGCACCTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCCATCGGCA 1902
QY 543 CCCCCTGCTGCCCGCACAAACATGTGGACCAAGTTCAGGAGTCTCAGGAGTACTCTGCAGCG 602
Db      |||||
1903 TCCACTGCTTCCCGCAGACACATGTGTTCTAGCCAGAGTTCAGGCCCACTCGACGA 1962
QY 603 CCAGGCGGTGGCGCGTGTGGGAGCGCTCATCTACCCGTCATGGAAGGCCATCGC 662
Db      |||||
1963 GATGGGTGCCCAATGCTTGGTCCACCATCATCTGTCGTCGATGATGATGAT 2022
QY 663 CCAGCCATGAAGTGGCGCCAGGACACATGAGCCCTCCAGAACAGCTTTGAGCTCTA 722
Db      |||||
2023 CCAGCAGCTTCAGACTCCCGCAGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTA 2082
QY 723 CGGGGCTGACTTCTGCTTCTGGAGGAGTTCAGGCGCTCGCTGATCGAGATCAATTCACAG 782
Db      |||||
2083 TGGCGCTGACTTCTGCTTCTGGGAGGAGTTCAGGCGCTCGCTGATGATGATGATGATGAT 2142
QY 783 CCCCACCATGCAACCGTCCACCGCGGTCAAGCCCGCTGCTGTGCAAGTGCAGGAGGA 842
Db      |||||
2143 CCCCACGATGGCACCTCCACAGCAGTCACTGCGCGGCTCTGTGCTGGCGTGCAGGCTGA 2202
QY 843 CACCATCAGTGGCGGTGGACCGCA 868
Db      |||||
2203 CACCCTGCGCGTGGTCAATGACCGGA 2228
```

```
RESULT 2
US-09-252-991A-2471
; Sequence 2471, Application US/09252991A
; Patent No. 6551795
; ORGANISM: Pseudomonas aeruginosa
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
```

```
; SEQ ID NO 2471
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2471

Query Match      5.8%; Score 52.6; DB 4; Length 1272;
Best Local Similarity 48.5%; Pred. No. 0.00059;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 177 GCACATTGACGGGCTCCGGAACATCTGGATTTATAAGCCCGCGGCAAGTCCCGGGCGG 236
Db      |||||
756 GGAATATGCTCGCGCAGCGGACAGGTACGGCAGATTTCCGGCAACCGCGAGGACGATCT 815
QY 237 AGACATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCC 296
Db      |||||
816 CGACGTACGGATGGTTCGACCAGCGCGCGCTGGAGGTGGTTCGAGTAGGCCAGGC 875
QY 297 TCTTTCCAGGGACAACAAAGTGGTGGTCCAGAAATGATCTGAGAGCGCGCTCTCATCTG 356
Db      |||||
876 GGTGACGCTGGAAGTCGAGTTCGAGTTCGCGCAGGACATCGAGCGGCTGATCTCGGCTT 935
QY 357 TGACACCAAGTTCGACATCAGACAGTGGTTCCTGTCAGGACTGGAAACCCCTGACCAT 416
Db      |||||
936 CATGATCAAGGACCGCCTCGGCCAGCGATGATCGGGATCAATACCCACCGCTGGACAA 995
QY 417 CTGGTTCTACAAAGGAGAGTTACTTTCGGTTCCTCAACTCAGCGCTTCTCCCTGGACAAGC 475
Db      |||||
996 GCGGCTACCGACCTGAAAGCGCGGAGCGGATCATCTACCGTTCGCTTCGACATGC 1054

RESULT 3
US-09-252-991A-2200/c
; Sequence 2200, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2200
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2200

Query Match      5.8%; Score 52.6; DB 4; Length 1449;
Best Local Similarity 48.5%; Pred. No. 0.00059;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 177 GCACATTGACGGGCTCCGGAACATCTGGATTTATAAGCCCGCGGCAAGTCCCGGGCGG 236
Db      |||||
562 GGAATATGCTCGCGCAGCGGACAGGTACGGCAGATTTCCGGCAACCGCGAGGACGATCT 503
QY 237 AGACATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCC 296
Db      |||||
502 CGACGTACGGATGGTTCGACCAGCGCGCGCTGGAGGTGGTTCGAGTAGGCCAGGC 443
QY 297 TCTTTCCAGGGACAACAAAGTGGTGGTCCAGAAATGATCTGAGAGCGCGCTCTCATCTG 356
Db      |||||
442 GGTGACGCTGGAAGTCGAGTTCGAGTTCGCGCAGGACATCGAGCGGCTGATCTCGGCTT 383
QY 357 TGACACCAAGTTCGACATCAGACAGTGGTTCCTGTCAGGACTGGAAACCCCTGACCAT 416
Db      |||||
382 CATGATCAAGGACCGCCTCGGCCAGCGGATGATCGGGATCAATACCCACCGCTGGACAA 323
```


QY 417 CTGTTCTACAGGAGAGTTACTTGGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 475
Db 322 GCGCTACCGACTGAAGCGCGGAGCGCATCACCTACCGCTTCGCTTCGACATGC 264

RESULT 4

US-09-252-991A-2632
; Sequence 2632 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2632
; LENGTH: 2790
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2632

Query Match 5.8%; Score 52.6; DB 4; Length 2790;
Best Local Similarity 48.5%; Pred. No. 0.00076;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 177 GGACATTCAGCGGCTCCGGACATCTCGATTATTAAGCCCGCCCAAGTCCCGGGCGG 236
Db 895 GGAATGCTCGCGACGCCAGGACGATTTCCGGCACCGCGGAGCGGATCCT 954
QY 237 AGACATAGTGTGACATGACCGTGTGGAGGAGATCCTGGAGCTGCAGCTGCAGACCAACC 296
Db 955 CGAGTACGATGTGCACACCGCGGAGCGGCTGGAGTGTGAGTAGGCGCAGGC 1014
QY 297 TCTTTCAGGAGCAACAAGTGGTGGTCCAGAGTACATCAGACGCGCTGTCTCATCTG 356
Db 1015 GGTGACGCTGGAAGTCGAGGTCGAGGTTCCGACGACATCAGCGGCTGTCTCGGCTT 1074
QY 357 TGACACCAAGTTCGACATCAGACAGTGGTTCCTGTCAGGACGGAACCCCTGACAT 416
Db 1075 CATGATCAAGGACCGCTCGCGCCAGCGATGTACGGGATCAATACCCACCGCTGGACAA 1134
QY 417 CTGTTCTACAGGAGAGTTACTTGGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 475
Db 1135 GCGCTACCGACCTGAAGCGCGGAGCGGATCACCTACCGCTTCGCTTCGACATGC 1193

RESULT 5

US-08-676-967-5
; Sequence 5, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-967-5

Query Match 5.4%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.0079;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 465 CCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAGCGCGTCCAGAACTACCTGAA 524
Db 1437 CGAAGACATCTTCTGTAGCGCGACCGCGCTGTGCCTGCACAACCTGCCAAGCGCGTGA 1496
QY 525 GAATGATGTGGCGCGAGCCCTCTGTGTCGCCGACACAACATGTGGACCAAGCAGGTT 584
Db 1497 CGACAAGCAGCTCGCAAGCTGTCTGTAGCGCGCACCCAGCGGCGAGAAGCGCGTGCAT 1556
QY 585 CCAAGGAGTACTGCAGCGCCAGCGCGCTGGCGCGCTGTGGGGGAGCGCTCATCTACCGTC 644
Db 1557 CAAGGAGTGCCTGTATGCGGACCTGAAGGGCGTGCACGCAACATGAAGGGCCAGAG 1616
QY 645 CATGAAGAAGGCCATCGCCCAACGCGCATGAAGGTGGCCAGGACCACTGTGGAGCTCGCAA 704
Db 1617 CTTGGGCTACGCTTCTGGCGAGTTCAGGAGCAGCAGCAGCGCTGAGGCCCTGCGCT 1676
QY 705 GAACAGC 711
Db 1677 GATCAAC 1683

RESULT 6

US-08-676-974-5
; Sequence 5, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-098-487-5

Query Match 5.4%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.0079;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 465 CTGGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAAGCGCGTCCAGAAGTACCTGAA 524
Db 1437 CCAGAACATCTTCGTGAGCGCGACCCGCTGTGCTGCAACCTGCCAAGCGCGTGA 1496

QY 525 GAATGATGTGGCGCGAGCGCCCTGCTGCTGCCCGCACACAACATGTGGACCAACCAAGTT 584
Db 1497 CGAACAGCAGCTGCGCAAGCTGTCTGAGCGCCACCGCGGAGAGCGGTGCGCAT 1556

QY 585 CAGGAGTACCTGCAGCGCGCAGCGCGTGGCGCGCGTGGCGCGCGTATCTTACCCGTC 644
Db 1557 CAGGAGTACCTGCAGCGCGCAGCGCGTGGCGCGCGTGGCGCGCGTATCTTACCCGTC 644

QY 645 CATGAAGAAGGCGCATCGCCACGCCATGAAGTGGCGCCAGGACCACTGCGAGCCTCGCAA 704
Db 1617 CTGGGCTACGCTTCGCGAGTTCAGGAGCAGGAGCAGCGCCCTGAAGCGCTTGGCCT 1676

QY 705 GAACAGC 711
Db 1677 GATCAAC 1683

RESULT 7
US-09-098-487-5
Sequence 5, Application US/09098487
Patent No. 5917025
GENERAL INFORMATION:
APPLICANT: COLLINS, Kathleen
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,487
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: cDNA
US-09-098-487-5

Query Match 5.4%; Score 48.6; DB 2; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.0079;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 465 CTGGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAAGCGCGTCCAGAAGTACCTGAA 524
Db 1437 CCAGAACATCTTCGTGAGCGCGACCCGCTGTGCTGCAACCTGCCAAGCGCGTGA 1496

QY 525 GAATGATGTGGCGCGAGCGCCCTGCTGCTGCCCGCACACAACATGTGGACCAACCAAGTT 584
Db 1497 CGAACAGCAGCTGCGCAAGCTGTCTGAGCGCCACCGCGGAGAGCGGTGCGCAT 1556

QY 585 CAGGAGTACCTGCAGCGCGCAGCGCGTGGCGCGCGTGGCGCGCGTATCTTACCCGTC 644
Db 1557 CAGGAGTACCTGCAGCGCGCAGCGCGTGGCGCGCGTGGCGCGCGTATCTTACCCGTC 644

QY 645 CATGAAGAAGGCGCATCGCCACGCCATGAAGTGGCGCCAGGACCACTGCGAGCCTCGCAA 704
Db 1617 CTGGGCTACGCTTCGCGAGTTCAGGAGCAGGAGCAGCGCCCTGAAGCGCTTGGCCT 1676

QY 705 GAACAGC 711
Db 1677 GATCAAC 1683

RESULT 8
US-09-471-033-18
Sequence 18, Application US/08471033
Patent No. 5770696
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..2004
OTHER INFORMATION: /note= "Maize optimized DNA
US-08-471-033-18
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"

Query Match 5.3%; Score 48.2; DB 1; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.0096;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 291 CCACCTCTTTCCAGGACCAACAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCT 350
DB 720 CCACCCCATCACCTGACAGAGAGAGTGGAGCACTGCTGAACCAAGCCCATGAT 779
QY 351 CATCTGTGACCAAGTTCGACATCGACAGTGGTTCCTCGTCAGGACTGGAAACCCCT 410
DB 780 GCTGGAGACCAACAGACCGCGCTCTACAAGATCAAGGACACCCACCGCAACATCGT 839
QY 411 GACCATCTGTTTACAGAGAGAGTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGA 470
DB 840 GACCGCGCGGAGTGGAAACCGCTGTATCCAGCAGATCAAGGCCAAGACCCCGCAGCATCAT 899
QY 471 CAAGCTGGACAGCGCCATCCACCTGTGCAACAACGCGCTCCAGAGTACCTGAAGATGA 530
DB 900 CGTCGACGACGGAGCGCGTGGCCGAGAGCGGTGGCGCCGAGACTACAGAACCC 959
QY 531 TGTGGGCGCAGCCCCCTGCTGCGCGCAACAACATGTGGACGACGACCCAGGTTCCAGGA 590
DB 960 CGAGGACAGACCCCGACCTGACCTGAGGAGCGCCCTGAAGTGAGCTACCCGACGA 1019
QY 591 GTACTGACGCGCAGCGCGCGTGGCGCGTGGGGAGCGGTATACCCGCTCATGA 650
DB 1020 GATCAAGGAGATCGAGGCGCTGCTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCGT 1079
QY 651 GAAGGCGCATCGCCACGCCATGAAGTGGCGCCAGGACCACTGGAGCTCGCAAGAACAG 710
DB 1080 GATGACCTATCTAGACGAGAACACCGCCAGAGAGGTGACCAAGAGCTGAACGACACCC 1139
QY 711 CTTTGAGCTTACGGGCTGACTTCGTCTTGGAGGGAATTTCAGGCGCTTGGCTGATCGA 770
DB 1140 CGGCAAGTTCAGGACGTGAGCCACCTGTACTAGACGTGAAGCTGACCCCGCAAGATGAAGT 1199
QY 771 GATCA 775
DB 1200 GACCA 1204

RESULT 9

US-08-471-044-18
Sequence 18, Application US/08471044
Patent No. 5840868
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Koshchka, N. Kristy
APPLICANT: Duck, Nicholas B

APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..2004
OTHER INFORMATION: /note= "Maize optimized DNA
US-08-471-044-18
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"

Query Match 5.3%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.0096;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 291 CCACCTCTTTCCAGGACCAACAGTGGTGGTCCAGAGTACATCGAGACGCGCTGCT 350
DB 720 CCACCCCATCACCTGACAGAGAGAGTGGAGCACTGCTGAACCAAGCCCATGAT 779
QY 351 CATCTGTGACCAAGTTCGACATCGACAGTGGTTCCTCGTCAGGACTGGAAACCCCT 410
DB 780 GCTGGAGACCAACAGACCGCGCTCTACAAGATCAAGGACACCCACCGCAACATCGT 839
QY 411 GACCATCTGTTTACAGAGAGAGTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGA 470
DB 840 GACCGCGCGGAGTGGAAACCGCGTGTATCCAGCAGATCAAGGCCAAGACCCCGCAGCATCAT 899
QY 471 CAAGCTGGACAGCGCCATCCACCTGTGCAACAACGCGCTCCAGAGTACCTGAAGATGA 530
DB 900 CGTCGACGACGGCGAGCGGTGGCCGAGAGAGCGGTGGCGCCGAGACTACAGAACCC 959

STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471.046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
US-08-471-046A-18

Query Match 5.3%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.0096;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 291 CCACCTCTTTCAGGGACAAAGTGGGTGGTCCAGAGTACATCGAGACGCGGTCT 350
DB 720 CCACCCCATCACCCTGACAGAGAGAGGTGGAGAACCTCTGACACCAAGCCATGAT 779
QY 351 CATCTGTGACCAAGTTCGACATCAGACAGTGGTTCCTGTCAGGACTGGAAACCCCT 410
DB 780 GCTGGAGACCAACAGACCGCGCTCAAGATCAAGGACACCCACCGCAACATCGT 839
QY 411 GACCATCTGTTCTACAGAGAGATTACTTGGGTCTTCACTCAGCGCTTCTCCCTGGA 470
DB 840 GACCGCGCGGAGTGGAAACCGCGTGTATCCAGCAGATCAAGCCCAAGACCGCCAGCATCAT 899
QY 471 CAAGCTGGACAGCGCCATCCACCTGTGTCAACAAACGCGCTCAGAGTACCTGAAGATGA 530
DB 900 CGTCGACGACGGAGCGCGCTGGCCGAGAGAGCGGTGGCCCGCCAGAGCTACAGAACCC 959
QY 531 TGTGGGCGGAGCCCTCTGTCGCCGCGACACAAATGTGGACCAAGCAGGTTCCAGGA 590
DB 960 CGAGGACAGACCCCGACCTGACCTTGAAGGAGCGCCTGAAGCTGAGCTACCCGACGA 1019

QY 591 GTACCTGACGCGCAGGCGCGTGGCGCGTGGGGCAGCGTCACTACCGTCCATGAA 650
DB 1020 GATCAAGGAGATCGAGGCGCTGCTGTACTACAAGAACAGCCCATCTACGAGACGCGT 1079
QY 651 GAAAGGCCATCGCCCAACGCCATGAAGGTGGCCCAAGGAGTGCACCAAGCAGCTGAACGACAC 710
DB 1080 GATGACCTATCTAGACGAGAAACACCGCAAGGAGGTGACCAAGCAGCTGAACGACAC 1139
QY 711 CTTTGAGCTCTACGGGCGTCACTTCTGTCCTTGGAGGAGACTTCAGGCGCTGGCTGATCGA 770
DB 1140 CGGCAAGTTTCAAGGACGTGAGCCACCTGTACGAGCTGAAGCTGACCCCAAGATGAACGT 1199
QY 771 GATCA 775
DB 1200 GACCA 1204

RESULT 12
US-08-470-566B-18
Sequence 18, Application US/08470566B
Patent No. 5872212
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

	TOPOLOGY:	linear	
	MOLECULE TYPE:	DNA (genomic)	
	HYPOTHETICAL:	NO	
	ANTI-SENSE:	NO	
	FEATURE:		
	NAME/KEY:	misc feature	
	LOCATION:	1..2004	
	OTHER INFORMATION:	/note= "Maize optimized DNA	
	OTHER INFORMATION:	sequence for VIP1A(a) 80 kd protein from AB78"	
	US-08-470-566B-18		
	Query Match	5.3%; Score 48.2; DB 2; Length 2004;	
	Best Local Similarity	43.7%; Pred. No. 0.0096;	
	Matches 212; Conservative	0; Mismatches 273; Indels 0; Gaps 0;	
Qy	291	CCACCCCTTTTCCAGGGACAACAAGTGGGTGTCAGAAGTACATCAGAGCGCGTGCT	350
Db	720	CCACCCCATCACCTGGAACAAGAAGCAGGTGGACAACCTGCTGAACAACAAGGCCATGAT	779
Qy	351	CATCTGTGCACCNAAGTTTCGACATCAGACAGTGGTTCCTCGTCACGCACTGGAAACCCCCT	410
Db	780	GCTGGAGACCAACAGACCCGACGGCGTCTACAAGATCAAGGACACCCACGGCAACATCGT	839
Qy	411	GACCATCTGGTCTTACAAGGAGAGTTACTTTCGGGTCTCAAATCAGCGCTTCTCCCTGGGA	470
Db	840	GACCGGCGCGAGTGGAAACGGCGTGATCCAGCAGATCAAGGCCAAGACCCGACGATCAT	899
Qy	471	CAGACTGGACAGGGCCATTCACCTGTGTGCACAACGCGCTGCAGAGTACTCTGAAGAATGA	530
Db	900	CGTCGACAGCGGGCAGCGCGTGGCCGAGAACGCGTGGCCGCCCAAGGACTACGAGAACCC	959
Qy	531	TGTGGGCGGAGCCCCCTGTGTCGCCGCACACAACATGTGGACCAAGCAGGTTCCAGGA	590
Db	960	CGAGGACAAGACCCCAAGCTGACCTTGAGGACGCCCTGAAGCTGAGCTACCCGACGA	1019
Qy	591	GTACTTCGACGCGCAGGGCCGTGGCGCGTGTGGGGCAGCGTCATCTACCCGTCCATGAA	650
Db	1020	GATCAAGGAGATCGAGGGCTGTGTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCGT	1079
Qy	651	GAAGGCCATCGCCACGCCATGAGGTGGSCCAGGACCAAGTGGAGCCTCGCAGAACAG	710
Db	1080	GATGACCTATTAGACGAGAACACCCGCAAGGAGGTGACCAAGCAGCTGAACGACCCAC	1139
Qy	711	CTTTGAGCTCTACGGGGCTGACTTCGTCTTGGGAGGAGCTTACGGGCCCTGGGTGATCGA	770
Db	1140	CGGCAAGTTCAAGGACGTGAGCCACCTGTGTACGACGTGAAGCTGACCCCAAGATGAACGT	1199
Qy	771	GATCA	775
Db	1200	GACCA	1204

```

RESULT 13
US-08-469-334-18
; Sequence 18, Application US/08469334
; Patent No. 5990383
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY

```

```

; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,334
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/463,483
; FILING DATE:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8615
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
;
US-08-469-334-18

Query Match 5.3%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.0096;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 291 CGACCTCTTTCAGGACACCAAGTGGTGTCCAGAGTACATCGAGACGGCGTGCT 350
   |||||
Db 720 CCACCCCATCACCTGAACAAGACGGGTGGAGCAACTCTGCTGAACAAAGCCCATGAT 779
   |||||

QY 351 CATCTGTGACCACAAGTTCGACATCAGACAGTGTCTTCCTCGTCAGGACTGGAAACCCCT 410
   |||||
Db 780 GCTTGGAGACCAACACAGCCAGCGGCTCTACAAGATCAAGACACCCACGGCAACATCGT 839
   |||||

QY 411 GACCATTCTGTCTTCAAGAGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGA 470
   |||||
Db 840 GACCGGGCGGCGAGTGGAACGGCGTGTATCCAGCAGATCAAGGCCAAGACCCGACGATCAT 899
   |||||

QY 471 CAAGCTGGACAGCGGCATCCACCTGTGTCAACACGCGGTCACAGAGTACCTGAGAGATGA 530
   |||||
Db 900 CGTCGACGACGGCGAGCGGTGGCCGAGAGCGGTGGCCGCCCAAGGACTACGAGAAACCC 959
   |||||

QY 531 TGTGGGCGCGAGCCCTCTGTGCGCCGCACACAACATGTGGACACGACCAGGTTCCAGGA 590
   |||||
Db 960 CGAGGACACAGACCCCGACCTGTGAGAGCGCCCTGAAGCTGAGCTACCCGACGA 1019
   |||||

QY 591 GTACTCTGACGCGCAGGGCGGTGGCCGCTGTGTGGGCAAGCGTCACTTACCCGTCATGAA 650
   |||||
Db 1020 GATCAAGGAGATCGAGGGCTGTGTACTACAAGAAACAAGCCATCTACGAGACGCGT 1079
   |||||

```


QY 651 GAAGGCGATCGCCACGCGATGAAGTGGCCAGGACACGCTGGAGCTCGCAAGAACAG 710
|||
Db 1080 GATGACCTATCTAGACGAGACACCGCCAGGAGGTGACCAAGCAGCTGAACGACACAC 1139
|||
QY 711 CTTTGAGCTCTACGGGGCTGACTTCGTCCTTTGGAGGAGCTTCAAGGCCCTGGCTGATCGA 770
|||
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
|||
QY 771 GATCA 775
|||
Db 1200 GACCA 1204
|||

RESULT 14

US-09-300-529-18
; Sequence 18, Application US/09300529
; Patent No. 6066783
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6066783artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,529
; FILING DATE: TBA
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/469,334
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-19506L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
US-09-300-529-18

Query Match 5.3%; Score 48.2; DB 3; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.0096;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 291 CCACCCCTCTTTCCAGGAGCAACAAGTGGGTGGTCCAGAACTACATCCAGAGCCCGCTGCT 350
|||
Db 720 CCACCCCATCACCTGAAACAAGAGCAGGTGGACAACTCTGTAACAACAAGCCCATGAT 779
|||
QY 351 CATCTGTGACACCAAGTTTCGACATCAGACAGTGTCTCTCGTCACGCGACTGGAAACCCCT 410
|||
Db 780 GCTGGAGACCAACAGACACCGCGCTTCAAGATCAAGGACACCCACCGCAACATCGT 839
|||
QY 411 GACCATCTGTTCTTACAAGGAGAGTTACTTTCGCGTTTCTCAACTCAGCGCTTCTCCCTGGA 470
|||
Db 840 GACCGCGCGAGTGGACCGCGTGTATCCAGCAGATCAAGGCCAAGACCCGCGACATCAT 899
|||
QY 471 CAAGCTGGACAGCGCCATTCACCTGTGTGCAACAACCGCGTCCAGAACTACCTGAAAGAATGA 530
|||
Db 900 CGTCGACGACGGCGCGCGCTGGCCGAGAGCGCGTGGCGCCCAAGGACTACGAGAACC 959
|||
QY 531 TGTGGCGCGAGCCCTCTGTCGCCGACACACATGTGGACCGACCGAGTTCCAGGA 590
|||
Db 960 CGAGGACAAGACCCCGCTGACCTTGAAGGACCGCTGAAGCTGAGCTACCCCGAGGA 1019
|||
QY 591 GTACCTGCAGCGCGCGCGTGGCGCGTGTGGGGCAGCGTTCATCTACCCGTCATGAA 650
|||
Db 1020 GATCAAGGAGATCGAGCGCTGTGTACTACAAGAACAAAGCCCATCTACGAGACAGCGT 1079
|||
QY 651 GAAGGCATCGCCACGCGCATGAAGTGGCGCCAGGACACGCTGGAGCCTCGCAAGAACAG 710
|||
Db 1080 GATGACCTATCTAGACGAGAACACCGCCCAAGAGGTGACCAAGCAGCTGAACGACAC 1139
|||
QY 711 CTTTGAGCTCTACGGGCTGACTTCGTCCTTTGGAGGAGCTTCAAGGCCCTGGCTGATCGA 770
|||
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
|||
QY 771 GATCA 775
|||
Db 1200 GACCA 1204
|||

RESULT 15

US-08-471-033-17
; Sequence 17, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA

ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2655 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..2652
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence for 100 kd V1P1A(a) protein from AB78"
US-08-471-033-17

Query Match 5.3%; Score 48.2; DB 1; Length 2655;
Best Local Similarity 43.7%; Pred. No. 0.011;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 291 CCACCCCTCTTCCAGGCAACAAGTGGGTGCTCCAGAGTACATCGAGACGCGCTGCT 350
DB 1371 CCACCCCATCACCCCTGAACAAGAGCAGGTGGACAACCTGCTGGAACAACAAGCCCATGAT 1430
QY 351 CATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCTCTCGTCACGGAAGTGAACCCCT 410
DB 1431 GCTGGAGACCAACACGACCGGCGCTCTACAAGATCAAGGACACCCACGCGAATCCT 1490
QY 411 GACCATCTGTTCTACAAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGA 470
DB 1491 GACCGCGCGAGTGAACGGCTGATCCAGCAGATCAAGGGCAAGACCGCCAGCATCAT 1550
QY 471 CAACTGGACAGCGCCATCCACTGTGCAACAACGCGCTCCAGAGTACTCTGAAGAATGA 530
DB 1551 CGTCGACGACGGCGAGCGGTGGCCGAGAGGCGGTGGCGGCAAGGACTACGAGAACCC 1610
QY 531 TGTGGCCCGCAGCCCTGCTGCCCCACACAACATGTGGAACGACGACCAAGGTTCCAGGA 590
DB 1611 CGAGGACAAGACCCCGAGCTGACCTCTGAAGGACGCCCTGAAGCTGAGCTACCCGACGA 1670
QY 591 GTACTGTGACGCGCAGGCGCGTGGCCGCTGTGGGGCAGCGTCACTACCGTCCATGAA 650
DB 1671 GATCAAGGAGATCGAGGGGCTGCTGTACTACAAGAAACAAGCCCATCTACGAGAGCAGCGT 1730
QY 651 GAAGGCCATCGCCCGCCATGAAGTGGCCCGGAGGACCGAGGAGCTTCGGAAGACAG 710
DB 1731 GATGACCTATCTAGACGAGAAACCCGCCAAGGAGGTGACCAAGCAGCTGAACGACACCAC 1790

QY 711 CTTTGAGCTCTACGGGGCTGACTTCGTCTGGGAGGAGACTTTCAGGCCCTGGCTGATCGA 770
DB 1791 CGGCAAGTTTCAAGGAGCTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1850
QY 771 GATCA 775
DB 1851 GACCA 1855

Search completed: September 24, 2005, 11:30:36
Job time : 166.553 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 11:08:28 ; Search time 3481.78 Seconds
(without alignments)
9904.766 Million cell updates/sec

Title: US-10-615-659-1_COPY_369_1274
Perfect score: 906
Sequence: 1 gagacatgcacgtcagc.....tcgagctctgtgagggcag 906

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_est3.*

4: gb_est4.*

5: gb_est5.*

6: gb_est6.*

7: gb_est7.*

8: gb_est8.*

9: gb_est9.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	906	100.0	1259	9	AY415398 Homo sapi
2	863.4	95.3	1257	9	AY415399 Pan trogl
3	768	84.8	770	7	CR745100 CR745100
4	608.8	67.2	1265	9	AY415400 Mus muscu
5	608.8	67.2	2720	3	AK029745 Mus muscu
6	608.8	67.2	2969	3	AK030151 Mus muscu
7	398.8	44.0	4184	3	EX648175 Homo sapi
8	378.4	41.8	1023	4	BM808637 AGNCOURT
9	377	41.6	2520	4	AK080321 Mus muscu
10	370	40.8	775	5	BU937372 AGNCOURT
11	360	39.7	1022	4	BM808516 AGNCOURT
12	320.6	35.4	693	4	BM717061 UI-E-EJ0-
13	316	34.9	861	7	CK653583 AGNCOURT
14	305.8	33.8	896	6	CD515431 AGNCOURT
15	294.6	32.5	777	7	CK597795 AGNCOURT
16	285.2	31.5	762	6	CD627148 56047432J
17	284.2	31.4	503	7	CK604006 AGNCOURT
18	257.8	28.5	583	7	CK349423 170004243
19	252.8	27.9	514	7	CR746623 CR746623
20	245.4	27.1	320	1	AA144756 mr73D04.r
21	243.4	26.9	680	6	CB854808 UI-CF-EC1
22	222.4	24.5	429	7	W31827 zb94f04.r1
23	219.2	24.2	880	5	EX448465 BX448465
24	213.2	23.5	616	7	CN349424 170006000

25	205	22.6	453	7	W31397
26	199	22.0	590	1	AL045121
c 27	195.6	21.6	484	2	BF830679
28	192.2	21.2	414	7	CN096245
29	191.4	21.1	875	5	BQ718874
30	190.8	21.1	3328	3	BC028169 Homo sapi
31	190.4	21.0	651	7	CF999072
c 32	186.2	20.6	504	5	BQ327883
33	185.6	20.5	965	5	EX431541
34	185.4	20.5	580	5	BP305902
35	179.8	19.8	724	4	BG916644
c 36	176.6	19.5	498	1	AA551103
37	173.6	19.2	488	9	CE760083
38	168.8	18.6	608	7	CN349425
39	167.6	18.5	812	6	CD648578
40	158.8	17.5	630	2	BE865183
41	156	17.2	850	5	EX849970
42	154	17.0	199	1	AA061487
43	151	16.7	507	2	BE837573
44	149.4	16.5	468	2	BE827753
45	146.6	16.2	977	5	EX431542

ALIGNMENTS

RESULT 1
LOCUS AY415398 1259 bp DNA linear GSS 12-DEC-2003
DEFINITION Homo sapiens HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION AY415398
VERSION AY415398.1 GI:39771357
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1259)
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1259)
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
source Location/Qualifiers
1..1259
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1259
/locus_tag="HCM5544"

ORIGIN
Query Match 100.0%; Score 906; DB 9; Length 1259;
Best Local Similarity 100.0%; Pred. No. 7e-208;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAGGACATGCACACGTGCGTGGAGCAGCTCACTGAGGCCGAGTGGAGGAC 60
|||||
234 GAGGACATGCACACGTGCGTGGAGCAGCTCACTGAGGCCGAGTGGAGGAC 293

QY	61	CTGACCCAGCAGTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT	120
Db	294	CTGACCCAGCAGTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT	353
QY	121	TACTTTTCGAGTGCAGAGCTCTGCTGAATAGAAATACAGTCTGTGAAACCTTCAGACGGAC	180
Db	354	TACTTTTCGAGTGCAGAGCTCTGCTGAATAGAAATACAGTCTGTGAAACCTTCAGACGGAC	413
QY	181	ATTGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGCCCAAGTCCCGGGGCCAGAC	240
Db	414	ATTGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGCCCAAGTCCCGGGGCCAGAC	473
QY	241	ATAGTGTGCATGCACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCCTCTT	300
Db	474	ATAGTGTGCATGCACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCCTCTT	533
QY	301	TCAGGGACAACAAAGTGGTGTCCAGAGTACATCGAGACGGCTGTCTCATCTGTGAC	360
Db	534	TCAGGGACAACAAAGTGGTGTCCAGAGTACATCGAGACGGCTGTCTCATCTGTGAC	593
QY	361	ACCAAGTTCGACATCAGACAGTGGTTTCCTCGTCAAGCTGGAAACCCCTGACCATCTGG	420
Db	594	ACCAAGTTCGACATCAGACAGTGGTTTCCTCGTCAAGCTGGAAACCCCTGACCATCTGG	480
QY	421	TTCTACAAGGAGTACTTTCGGTTTCTCAACTCAGCGCTTCTCCTTGGACAAGCTGGAC	480
Db	654	TTCTACAAGGAGTACTTTCGGTTTCTCAACTCAGCGCTTCTCCTTGGACAAGCTGGAC	713
QY	481	AGGGCCATCAGCTGTGCAACAAACCGCTGCAGAGTACCTGAAGATCATGTGGGCCCGC	540
Db	714	AGGGCCATCAGCTGTGCAACAAACCGCTGCAGAGTACCTGAAGATCATGTGGGCCCGC	773
QY	541	AGCCCCCTCTCTCCGCGCACACAACATGTGGACCAAGCTTCCAGAGTACCTTGCAG	600
Db	774	AGCCCCCTCTCTCCGCGCACACAACATGTGGACCAAGCTTCCAGAGTACCTTGCAG	833
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.			
1 (bases 1 to 1257)			
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,			
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,			
Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,			
Adams, M.D. and Cargill, M.			
Inferring nonneutral evolution from human-chimp-mouse orthologous			
gene trios			
Science 302 (5652), 1960-1963 (2003)			
14671302			
2 (bases 1 to 1257)			
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,			
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,			
Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,			
Adams, M.D. and Cargill, M.			
Direct Submission			
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,			
Rockville, MD 20850, USA			
This sequence was made by sequencing genomic exons and ordering			
them based on alignment.			
Location/Qualifiers			
Source			
1..1257			
/organism="Pan troglodytes"			
/mol_type="genomic DNA"			
/db_xref="taxon:9598"			
<1..>1257			
/locus_tag="HCM5544"			
Query Match 95.3%; Score 863.4; DB 9; Length 1257;			
Best Local Similarity 95.9%; Pred. No. 1.3e-197;			
Matches 867; Conservative 0; Mismatches 37; Indels 0; Gaps 0;			
QY	1	GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAC	60
Db	234	GAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAC	293
QY	61	CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT	120
Db	294	CTGACCCAGCAGTACTCTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAAGAAAT	353
QY	121	TACTTTTCGAGTGCAGAGCTCTGTGAAATAGAAATCAAGTCTGTGAAACCTTCAGACGGAC	180
Db	354	TACTTTTCGAGTGCAGAGCTCTGTGAAATAGAAATCAAGTCTGTGAAACCTTCAGACGGAC	413
QY	181	ATTGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGCCCAAGTCCCGGGGCCAGAC	240
Db	414	ATTGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGCCCAAGTCCCGGGGCCAGAC	473
QY	241	ATAGTGTGCATGCACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCCTCTT	300
Db	474	ATAGTGTGCATGCACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCCTCTT	533
QY	301	TCAGGGACAACAAAGTGGTGTCCAGAGTACATCGAGACGGCTGTCTCATCTGTGAC	360
Db	534	TCAGGGACAACAAAGTGGTGTCCAGAGTACATCGAGACGGCTGTCTCATCTGTGAC	593
QY	361	ACCAAGTTCGACATCAGACAGTGGTTTCCTCGTCAAGCTGGAAACCCCTGACCATCTGG	420
Db	594	ACCAAGTTCGACATCAGACAGTGGTTTCCTCGTCAAGCTGGAAACCCCTGACCATCTGG	480
QY	421	TTCTACAAGGAGTACTTTCGGTTTCTCAACTCAGCGCTTCTCCTTGGACAAGCTGGAC	480
Db	654	TTCTACAAGGAGTACTTTCGGTTTCTCAACTCAGCGCTTCTCCTTGGACAAGCTGGAC	713
QY	481	AGGGCCATCAGCTGTGCAACAAACCGCTGCAGAGTACCTGAAGATCATGTGGGCCCGC	540
Db	714	AGGGCCATCAGCTGTGCAACAAACCGCTGCAGAGTACCTGAAGATCATGTGGGCCCGC	773
QY	541	AGCCCCCTCTCTCCGCGCACACAACATGTGGACCAAGCTTCCAGAGTACCTTGCAG	600
Db	774	AGCCCCCTCTCTCCGCGCACACAACATGTGGACCAAGCTTCCAGAGTACCTTGCAG	833

RESULT 2
AY415399
LOCUS
DEFINITION
AY415399
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AY415399 1257 bp DNA linear GSS 12-DEC-2003
Pan troglodytes HCM5544 gene, VIRUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
AY415399
AY415399.1 GI:39771358
GSS.
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1265)
REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source
1..1265
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
gene
1..1265
/locus_tag="HCM5544"
ORIGIN
Query Match 67.2%; Score 608.8; DB 9; Length 1265;
Best Local Similarity 80.0%; Pred. No. 4e-136;
Matches 734; Conservative 0; Mismatches 172; Indels 12; Gaps 1;
QY 1 GAGGACATGCACACGTCAGACATGCGTGGAGGACTCCTGAGGCGGAGGAGGAC 60
DB 228 GAAGACATGAGATATCGGAGGCGGAGCAGCGAGGCTTGTCTGAGGAGGAATGATGAC 287
QY 61 CTGACCCAGCAGTACTCTCCCTCGTTTCATGGGGATGCTTTTCATCTCCAAATCAAGAAAT 120
DB 288 CTGACACACAGTACTATCTACTGTTTCATGGCAATGCTTCCATCACTGATTCGAAAAGT 347
QY 121 TACTTTTCGAGTGCAGAGGCTCTGCTGAATAGAAATCAGCTCTGTGAACCTCTGACGGAC 180
DB 348 TACTTTTCGCAATGCCAAGCTCTGCTGAGTAAGATCAGTCTGTGAAACCTCTCAGACAGAG 407
QY 181 ATTGACGGCTCCGGAACATCTGGATTATAAAGCCGCGGCAAGTCCCGGGGCCGAGAC 240
DB 408 ATCGATGGGATCCGGAACATCTGATCATTAAGCTTCGACGCGAAGTCCCGGGTCCGAT 467
QY 241 ATAGTGTGATGACCGCTGTGGAGGATCTCTGGAGCTGGCAGCTGCAGACCACTCTTT 300
DB 468 ATTGTGTGTGATGACCGTGTGGAGACATCTCTGAGTCTGTGGCGGCGACAGCCAGACC 527
QY 301 TCCAGGCAACAAGTGGGTGTCAGAGATCATCGACAGCGCGGTGCTCATCTGTGAC 360
DB 528 ACAAGGCAACAATAAGTGGTGTCCAGAAATCATCGACAGCCCAATGCTCATCTACGAC 587
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGCTGGAAACCCCTGACCATCTCG 420
DB 588 ACCAAGTTTGACATCAGACAGTGGTTCCTCGTCAAGCTGGAAATCCCTTAACCATCTGG 647
QY 421 TTCTACAAGGAGAGTGTACTTTGCGGTTTCTCAACTCAGCGTTTCTCCCTGGACAAGCTGGAC 480
DB 648 TTCTACAAGAAAGCTACCTCCGCTTCTCCACACAGGGCTTCTCCCTGGACAAGCTGGAC 707
QY 481 AGGCGCATCAGCTGTGCAACAACCGGCTCCAGAAGTACCTGAAGATGATGTGGGCCCGC 540
DB 708 AGGCGCATCAGCTGTGTAACAACCTCCATCCAGAGGCGCTCAAAAATGATTAAGAGCGC 767
QY 541 AGGCGCCCTGTCGCCGACACACATGTTGGACCGACGAGGTTCCAGGAGTACCTGCGAC 600
DB 768 AGTCCGCTGCTACTTGTCTATAACATGTGGACGAGCACTGCTTTCCAGGAGTACCTGCGAC 827

QY 601 COCCAGGGCCGTGGCGCCCGTGTGGGCGACGCTCATCTACCGTCCATGAGAAGGCATC 660
DB 828 AAGAGGGGCGGAGGAGAACGTTGGGGTAGCATCATCTACCCGCTCTATGAAAAGAGCTGTC 887
QY 661 GCCCAGCCCATGAAGTGGCCCGGAGGACCGTGGAGCCTCGCAAGACAGCTTTGAGCTC 720
DB 888 ACCAATGCCATGAGGTAGCCCGAGGACCGTAGAAGCCCGTAGAACACAGCTTCGAGCTC 947
QY 721 TACGGGCTGACTTCGTCTCTGGAGGAGACTTCAGGCCCTGGCTGATCGAGATCAATTC 780
DB 948 TACGAGCGGACTTCATCTCTGGGCGAGACTTCAGGCCATGGCTATCGAGATCACTCC 1007
QY 781 AGCCCCCACCATGACCCGTCACGCGGTCAGCGGCCAGCTGTGTGCA CAGGTGCAAGGAG 840
DB 1008 AGCCCCCACCATGACCCCTCCACTCCCGTCACAGCTCAGCTCTGTGCCAGGTGCAAGGAG 1067
QY 841 GACACCATCAAGTGGCGG-----TGGACCGCAGCTGTGACATCGGCAACTTC 888
DB 1068 GACACCATCAAGTGGTGTGTGATCGCAAACTGGACCGAACTGTGACATCGGCAACTTT 1127
QY 889 GAGCTCCTGTGGAGGCGAG 906
DB 1128 GAGCTCTCTGTGGCGGAG 1145
RESULT 5
AK029745 2720 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched
DEFINITION library, clone:490524K07 product:hypothetical Tubulin-tyrosine
ligase containing protein, full insert sequence.
ACCESSION AK029745
VERSION AK029745.1 GI:26325681
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL 99279253
MEDLINE 10349636
PUBMED
REFERENCE
2 Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL 20499374
MEDLINE
PUBMED 11042159
REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL 20530913
MEDLINE
PUBMED 11076861
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL
MEDLINE
PUBMED
REFERENCE
5 The FANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

TITLE

JOURNAL

REFERENCE

AUTHORS

6 (bases 1 to 2720)
Nature 420, 563-573 (2002)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayateu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Segabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp).
URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222,
Fax: 81-45-503-9216

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>

URL: <http://fantom.gsc.riken.jp/>.

FEATURES

source

1. 2720
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM:DB:4930524K07"
/db_xref="taxon:10090"
/clone="4930524K07"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
934..2490

CDS

/note="unnamed protein product; hypothetical
Tubulin-tyrosine ligase containing protein
(InterPro|IPR004344, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAC26595.1"
/db_xref="GI:26325682"
/translation="MLNHVYKSTAFSTKIGLCLNMRSLPWVQANPTFPFCVGLCT
ESKQBLDPFRVTAASILKVLVHONYKSVKGSKEEAKNSDPSPKDPEPDL
KLPSLQVDVTACKVQAYLGLEHEDIDVSEASTALESEENDLTQYLLVHGN
ASITDSKSYAQCALLKSSVNPQTEIDGIRNIWIIPKPAKSRGDIIVCMRDVENI
LSLVADSSQTKNKKVQYIETPMLIYDTDFIQWFLTDWNPDLTIWFKESYLK
FSTQFSLDLKLSAHLKCNNSIQRLKNDKERSPLPCHNWTSTRFQYLRKRGSG
TWGSIIVPSMKRAVTNMRVADQHVAKNSFELYGADFILGRDFKPLWLEINSPTM
HPSPVTAQLCAQVQETIKVVDRLDRNCIDIGNPELLWQPAVELPPFNGSDLCVE
GISVKKAKQMPPIASVGLSESLLDAPPKVRSAALMETVIRPRTTVRDQWRBEAK
GFRVISKRRR"

ORIGIN

Query Match 67.2%; Score 608.8; DB 3; Length 2720;
Best Local Similarity 80.0%; Pred No. 4.5e-136;
Matches 734; Conservative 0; Mismatches 172; Indels 12; Gaps 1;
Qy 1 GAGGACATCGACACGTGAGGAGTCCGCTGGAGACCTCACTGAGGCGGAGTGGAGGAC 60
Db 1318 GAAGACATAGACGTATCGAGGCCGACCGAGGCGCTTGCTGAGGAGGATGGAATGAC 1377

Qy 61 CTGACCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAATTCAGAAAT 120
Db 1378 CTGACAGCAGTACTACTTACTGGTTTCATGGCAATGCTTCCATCATGATTGCAAAAGT 1437
Qy 121 TACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCAAGTCTGTGGAACCTTCAGACGAC 180
Db 1438 TACTTTGCGCAATGCGCAAGCTCTGCTGAGTAAGATCAGTTCTGTGAAACCTTCAGACAG 1497
Qy 181 ATTGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGGCAAGTCCCGGGCGGAGAC 240
Db 1498 ATCGATGGGATCCGGAACATCTGGATCATTAAGCTTCAGACCAAGTCCCGGGCTCGAGAT 1557
Qy 241 ATAGTGTGATGACACCGCTGTGGAGGACATCTGAGCTGCGAGCTGCAGACCACTCTT 300
Db 1558 ATTGTGTATGACCGCTGTGGAGAACATCTGAGTCTGGTGGCCGAGACAGCAGAC 1617
Qy 301 TCCAGGACCAACAGTGGTGGTCCAGAACTACATCGAGACGCGCTGCTCATCTGTGAC 360
Db 1618 ACAAGGACCAACAAATGGTGGTCCAGAACTACATCGAGACCGCCCAATGCTCATCTAGAC 1677
Qy 361 ACAAGTTCGACATCAGACAGTGGTCTCGTACGAGCTGGAACCCCTGACCATCTGG 420
Db 1678 ACCAAGTTTGACATCAGACAGTGGTCTCGTACGAGCTGGAATCCCTTAACCATCTGG 1737
Qy 421 TTCTACAAGCAGAGTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACCAAGCTGAC 480
Db 1738 TTCTACAAAGAAAGTACTTCCGCTTCTCACAACAGGCTTCTCCCTGGACCAAGCTGAC 1797
Qy 481 AGCGCCATCCACCTGTGTCACAAACCGCTCCAGAACTACCTGAAGAAATGATGTGGCGGC 540
Db 1798 AGCGCCATCCACCTGTGTAACAACTCCATCCAGAGGCGCTCAAAAATGATAAGAGCGC 1857
Qy 541 AGCCCCCTGTGCGCGGACACAACATGTGACACGACACCAAGTTCCAGGAGTACCTGAG 600
Db 1858 AGTCGCTGTCTGCTTGTCTATAACATGTGACCAAGCTGCTTTTCCAGGAGTACCTGAG 1917
Qy 601 CGCCAGCGCGTGGCGCGTGTGGCGAGCTCATCTACCGCTCCATGAAGAGAGGCCATC 660
Db 1918 AAGAGGCGCGAGGAGGAGAACGTGGGTAGCATCATCTACCGCTCATGAAGAGAGCTGTC 1977
Qy 661 GCCACGCCATGAAGGTGGCGCCAGGACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTC 720
Db 1978 ACCAATGCCATGAGGTAGCCAGGACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTC 2037
Qy 721 TAGCGGCTGACTTGTCTTGGAGGAGACTTCAGGCGCTTGGCTGATCGAGATCAATTC 780
Db 2038 TAGCGAGCGGACTTCTATCTGCGGCGAGACTTCAAGCCATGGCTTATCGAGATCAATTC 2097
Qy 781 AGCCCCACCATGACCGCTCCAGCGGTCACGCGCCAGCTGTGTCAGAGTGCAGGAG 840
Db 2098 AGCCCCACCATGACCGCTCCAGCTCCCGTCCAGCTCAGCTCTGTGCGCCAGGTCAGGAG 2157
Qy 841 GACACCATCAAGTGGCGCG-----TGACCGCAGCTGTGACATCGGCAATTC 888
Db 2158 GACACCATCAAGTGGTGGTGGATCGCAAACTGACCGCAAACTGTGACATCGGCAATTT 2217
Qy 889 GAGCTCTGTGGAGGACG 906
Db 2218 GAGCTCTGTGGCGGACG 2235
RESULT 6
AK030151
LOCUS
DEFINITION
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4933401B17 product: hypothetical Tubulin-tyrosine
ligase containing protein, full insert sequence.
AK030151
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kicsuna, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 2969)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
Location/Qualifiers
1. .2969
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"

REFERENCE source

/db_xref="FANTOM DB:4933401B17"
/db_xref="taxon:10090"
/clone="4933401B17"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
490. .2835

CDS

note="unnamed protein product; hypothetical
Tubulin-tyrosine ligase containing protein
(InterPro: IPR004344, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAC26811.1"
/db_xref="GI:26326135"
/translation="MENGKKLSTLSGDHKEENKLGKQIGPQDLSSSPKLDRYKIA
RQLEKAIKERKIESYGHYPVIRLATRKGVVEKFFPKALQNLGSEKSAETKE
NOEIALERDDLDHDMRLVKNIEPYLLWTIKRDVVDVHSLTCDQMLNHYGKTASTTT
KIGLCNMRSLPWYVOANPNTFFPCYGLPCTSEKQEFLEDDFRRTVAASILKXVLUHQ
NYCSVKGSKKEEAKNDSPSPKPNPDLKPSLGQVVDPTACKVCOAYLGQLEHE
DIDVEASTALSEBWNLDLTOYLLVHGNASITDSKSYFAQCALLSKISVNPQT
EIDGIRNTWIIKPAASKRGDIVCDRVENILSLVAADSOTTCKDNKWWQVKYTEML
IYDTRKDIQWFLVTDWNPLOTIFKESYLRESTORFSLDKLSDAHLNCSIORLKL
KNKERSPLPCNMMTSTRPQYLOKRGRTGWSIILPMSKBAVTNMRVAQDHVEA
RNSFELYGADFILGRDPLFELIENSSPTFPAQLCAKQAVNOEDTIKVVVDRL
DKRCDIGNFELLWRQPAVELPFGNSDLCEVIGSVKAKKQMPPTAEVRSLESLDAP
PRVSGARALMETVIRPRTTVRQDWKREAKVLSTWSPMVDAAEVRGAKETIYAFEV
NDYQHDNKHSGSVTRVQSSKVPGLTSAQHPALFAQTMKPTQWTSPPPTASGNH
RDSFPFCPIVFEELMLHFNPSRRPSSILQSRAGQWIRGIP"

ORIGIN

Query Match 67.2%; Score 608.8; DB 3; Length 2969;
Best Local Similarity 80.0%; Pred. No. 4.6e-136;
Matches 734; Conservative 0; Mismatches 172; Indels 12; Gaps 1;

QY 1 GAGGACATCGACAGTCAGCAGATGCGCTGGGAGGACCTCACTGAGCCGAGTGGGAGGAC 60
DB 1315 GAAGACATAGACGTATCGGAGGCCAGCACCGAGGCTTGTCTGAGGAGGAATGGAATGAC 1374
QY 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAGAAAT 120
DB 1375 CTGACACAGCAGTACTACTCTGTTTCATGGCAATGCTTCCATCTACTGATTCGAAAAGT 1434
QY 121 TACTTTTCGAGTGGCAGGCTCTGCTGAATAGAATCACTGTGTGAACCTCAGACGGAC 180
DB 1435 TACTTTGCGCAATGCCAAGCTCTGCTGAGTAAGATCAGTTCTGTGAACCTCAGACAGAG 1494
QY 181 ATTGACGGGCTCGGACACATCTGGATTATTAAGCCGCGGCGCAAGTCCCGGGCCGAGAC 240
DB 1495 ATCGATGGGATCGGAACATCTGGATCATATAAGCCCTGACGCAAGTCCCGGGGTCGAGAT 1554
QY 241 ATAGTGTGCATGACCGCTGTGGAGAGATCTCTGGAGCTGCGAGCTGCAGACCAACCTCTT 300
DB 1555 ATTGTTGTATGGACCGTGTGGAGAACATCTGTAGTCTGGTGGCCGACAGACCCAGACC 1614
QY 301 TCCAGGGACAAACAGTGGTGGTTCGAGAGTAGTACATCGAGACCGCGCTGCTCATCTGTGAC 360
DB 1615 ACAAGGACAAACAAATGGGTGGTTCGAGAGTAGTACATCGAGACCCCAATGCTCATCTACGAC 1674
QY 361 ACCAAGTTGCATCATCAGACAGTGGTTCCTCGTCAAGTACCGGACCCCTGACCATCTGG 420
DB 1675 ACCAAGTTTGATCATCAGACAGTGGTTCCTCGTCAAGTACCGGACCCCTGACCATCTGG 1734
QY 421 TTCTCAAGGAGAGTACTTTCGGGTTCTCAACTCAGCGCTTCTCCTCGTCAAGTACCGTGGAC 480
DB 1735 TTCTCAAGGAGAGTACTCTCGCTTCTCCAACAGCGGTTTCTCCTCGACAGTGGAC 1794
QY 481 AGCGCCATCCACCTGTGCAACAAACCGCTGCAGAGTAGTACCTGAAGAATGATGTGGGCCGC 540
DB 1795 AGCGCCATCCACCTGTGTAAACAACTCCATCCAGAGCGCGCTCAAAAATGATAAGAGCGC 1854
QY 541 AGCCCCCTGCTCCCGCACACAAATGTGGACAGCACAGGTTCCAGAGTAGTACCTGCGAG 600


```
Db 1855 AGTCGGCTGTACCTTGTCTATACATGTGGACACAGCACTCGTTTCCAGGAGTACCTGCAG 1914
Qy 601 CGCCAGGGCGGTGGCGCGTGTGGGCGAGCGTCTATCTACCGCTCATGAGAACGCCATC 660
Db 1915 AAGAGGGGCGGAGGAGAACGTTGGGGTAGCATCTATCTACCGCTCTATGAAAGAGCTGTC 1974
Qy 661 GCCACGCGCATGAAGGTGGCGGAGCAGCACGCTGGAGCCTCCAGAACAGCTTTGAGCTC 720
Db 1975 ACCAATGCCATGAGGTAGCCAGGACCACTAGAGCCGCTAGAACAGCTTCGAGCTC 2034
Qy 721 TACGGGCGTGAATTCGCTTCTTGGAGGAGGACTTCAGGCGCTCGCTGATCGAGATCAATTC 780
Db 2035 TACGGAGCCGACTTTCATCTCTGGGCGAGACTTCAAGCCATGGCTTATCGAGATCAATCC 2094
Qy 781 AGCCCGACCATGACCCGCTCACCGCGGTACGCGCCAGCTGTGTGACAGGTGCAGAG 840
Db 2095 AGCCCCACCATGCAACCCCTCCACTCCCGTCAAGCTCAGCTCTGTGCCAGGTGCAGGAG 2154
Qy 841 GACACCATCAAGGTGGCGG-----TGGACCGGAGCTGTGACATCGGCAACTTC 888
Db 2155 GACACCATCAAGGTGGTGGTGTGATGCGCAAACTGGACCGAACTGTGACATCGGCACTTT 2214
Qy 889 GAGCTCTCTGGAGGCGAG 906
Db 2215 GAGCTTCTGTGGCGGCGAG 2232

RESULT 7
HSM808322 4184 bp mRNA linear HTC 22-SEP-2004
LOCUS Homo sapiens mRNA; cDNA DKFZp686D076 (from clone DKFZp686D076).
ACCESSION BX648175
VERSION BX648175.1 GI:34367334
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4184)
AUTHORS Ansoorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
Mewes,H.W., Weil,B., Amid,C., Oeinger,A., Fobo,G., Han,M. and
Wiemann,S.
CONSTRM The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686D076) is available at the RZPD Deutsches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686D076
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.

FEATURES
source
1..4184
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp686D076Q"
/db_xref="taxon:9606"
/clone="DKFZp686D076"
/tissue_type="cervix"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
DH10B; sites SfiI + SfiIb"
/dev_stage="adult"
/notes="hypothetical protein"
1..4184
/gene="DKFZp686D076"
2529..3587
/cds
/gene="DKFZp686D076"
```

```
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH10554.1"
/db_xref="GI:50949619"
/translation="MDIDKDLLEAPLYLTTPGWSLFLQRYQYVHGEALRHLDTVQVR
CEILQQLQAVVPQIDMEGRNTWIKPGAKSRGIMCMMDHLEMLKLVNGNPNVWK
DGKVVQKYIERPLLIIGTKFDLROWPLVTDNPLTVWFYRDSVIRESTOPFSUKNLD
NSVHLNNSLOKHELENSCHRPILLPDPNWSORFOAHLQEMGAPNAWSITLIVPMKD
AVTHALQTSQDVTQCRKASFELYGADFVFGEDFPWLEINASNATPASTATVATRLCA
GQADTLRVVVDRLDRNCDTGAFELIYKQPVTTSPASTPGPSPCLLPYSDTRASSD
DSTAWSWALRPCRQARP"

ORIGIN
Query Match 44.08; Score 398.8; DB 3; Length 4184;
Best Local Similarity 66.3%; Pred. No. 2.4e-85;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

Qy 3 GGACATCGACACGTCAGCAGATCCGCTGGAGGACCTCACTGAGCGCCAGTGGAGGACCT 62
Db 2531 GGACATCGACAGGACCTGGAGGCGCCGCTGTACTCACCCCGAGGGCTGCTCCTCTT 2590
Qy 63 GACCCAGCAGTACTACTCCTCTCATGCGCATGCTTTTCATCTCCAAATCAAGAAATTA 122
Db 2591 CCTCCAGCGCTACTACCAAGTGGTCCACGAAGGGGAGAACTCAGGCACCTCGACACTCA 2650
Qy 123 CTTTTCCAGTGCAGGCTCTGCTGATAGATACAGTCTGTGAACCTCAGACGGACAT 182
Db 2651 GGTCCAGCGCTGTGAGGACATCTCTGCAGCAGCTGCAGGCGCTGGTATCCCAAGATAGACAT 2710
Qy 183 TGACGGGCTCCGGAACATCTGGATATTAAAGCCGCGGCCAAAGTCCCGGGCCGAGACAT 242
Db 2711 GGAAGGGGATCGCAACATCTGGATCGTGAAGCAGGAGCCAGTCCCGCGGACAGGACAT 2770
Qy 243 AGTGTGCATGAGCCGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTTC 302
Db 2771 CATGTGCATGGACCACTCGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGTAT 2830
Qy 303 CAGGGAACAAAGTGGTGTCTCAGAAGTATCATCGACACCGCTGCTCATCTGTGACAC 362
Db 2831 GAAGGACGGCAAGTGGTGGTGCAGAGTATATTTAGCGCGGCCCTCTCTCATCTTTGGCAC 2890
Qy 363 CAAGTTTCAGATCAGACAGTGGTTCCTCGTCAAGCTGCAACCCCTGACCATCTCTGTT 422
Db 2891 CAAGTTTGAATCAGACAGTGGTTCCTGCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2950
Qy 423 CTCAAAGGAGATTAATCTGCGGTTCTCAACTCAGCGCTCTCTCCCTGGAAGCTGGAAG 482
Db 2951 CTACCGGACAGCTATATACCGCTTTTCCAGCAGCGCTCTCTCCCTGGAAGTGAAGTGAAGTGAAGT 3010
Qy 483 CGCATCCACTGTGCAACCAACCGCTCAGAGTACCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 542
Db 3011 CTCAGTGACACTGTGCAACCAACTCCATCCAGAAAGCACTGGAGAACTCATGCCATCCGCCA 3070
Qy 543 CCCCTGCTGCGCGCACACAACTGTGGACAGCAGCAGCTTCCAGGAGTACCTGCGAGCG 602
Db 3071 TCCACTGCTTCCGCGCAGAACAACTGCTGCTAGCAGAGGTTCAGGCGCCACCTGCGAGGA 3130
Qy 603 CCAGGCGCTGGCGCGCTGTGGGCGCAGCTCATCTACCCCTCATGAAGAGGCGCATCGC 662
Db 3131 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCGTGGCATGAAGGATGCTGTGAT 3190
Qy 663 CCACGCGCATGAAGGTGGCCCAAGCAACGCTGGAGCCTTCGCAAGAACAGCTTTGAGCTCTA 722
Db 3191 CCACGCACTTCAGACCTCCCGGACACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTA 3250
Qy 723 CGGGGCTGACTTCTGCTTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTCAG 782
Db 3251 TGGCGCTGACTTCTGCTTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAACGCGCAG 3310
Qy 783 CCCACCATGACCCCGTCCACGCGGTCCAGGCGCCAGCTGTGTGACAGGTGCAGGAGGA 842
Db 3311 CCCACCATGACCCCGTCCACGAGTCACTGCTGCGCGCTCTGTGCTGGCGTGCAGAGCTGA 3370
```


QY 843 CACCATCAAGTGGCGCTGGACCGCA 868
|||||
Db 3371 CACCCTGGCGTGGTCAATGACCGGA 3396
5', mRNA sequence.

BM808637 1023 bp mRNA linear EST 05-MAR-2002
AGENCOURT 6617786 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734232
5', mRNA sequence.

BM808637
BM808637.1 GI:19125460
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1023)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-f@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12739 row: d column: 09
High quality sequence start: 6
High quality sequence stop: 664.
Location/Qualifiers
1..1023
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734232"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH MGC 124"
/notes="Organ: Brain; Vector: pCMV-SPORT6; Site_1: EcoRV
(destroyed); Site_2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."

ORIGIN
Query Match 41.8%; Score 378.4; DB 4; Length 1023;
Best Local Similarity 69.5%; Pred. No. 1.7e-80;
Matches 514; Conservative 0; Mismatches 226; Indels 0; Gaps 0;
QY 129 GCAGTCCAGGCTCTGTAATAGATCAAGTCTGTGAACCCCTCAGACGGACATTGACGG 188
Db 49 GCCTGTGAGGACATCTCGACGAGCTGACGGCGGTGTACCCCATAGATAGATGGAAG 108
QY 189 GCTCCGGAACATCTGGATTATTAAGCCCGGCCCAAGTCCCGGGCCGAGACATAGTGTG 248
Db 109 GGATCGCAACATCTGGATCTGTGAAGCCAGGAGCCAAAGTCCCGCGGACGAGGCATCATG 168
QY 249 CATGGAACCTGTGGAGAGATCTCGAGCTGGAGCTGCAGACCCCTCTTTCAGGGA 308
Db 169 CATGGACCACTTGAGAGATGCTGAAGCTGTGTGAACGGCAACCCCGTGTGATGAAGGA 228
QY 309 CAACAAGTGGTGGTCCAGAAGTACATCGAGACCGCGCTCTCATCTGTGACACCAAGTT 368
Db 229 CGGCAAGTGGTGGTGCAGAGATATATTGAGCGGCCCTCTCATCTTTGGCACCAAGTT 288
QY 369 CGACATCAGACAGTGGTCTCTCGTCA CGGACTGGAA CCCCCTGACCATCTGGTTCTACAA 428

Db 289 TGACCTCAGACAGTGGTCTCTGTAACCTGACTGGAACCCACTTACCGTGTGGTTTACCG 348
QY 429 GGAGAGTTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTCGACAGCCCAT 488
Db 349 CGACAGCTATATCCGCTTTTTCACGCGAGCCCTTCTCCCTGAAAGAACTCTGGACAACTCAGT 408
QY 489 CCACTGTGTCAACACCGCGTCCAGAAATACCTGAAGAAATGATGTGGCGCGCAGCCCCCT 548
Db 409 GCACCTGTGCAACNACTCCATCCAGAAGCACCTGGAGAACTCATGCCATCGGCATCCACT 468
QY 549 GCTGCCCGCACACACATGTGACACAGCACCGAGTTCCAGAGTACCTGAGCGCCAGGG 608
Db 469 GCTTCGCCCAGACACATGTGGTGTAGCCAGAGGTTTCCAGGCCCCACCTCGAGGAGATGG 528
QY 609 CGGTGGCGCGGTGGGGCAGCGTCTATCTACCCGCTCATGAAGAGGCCATCGCCACCGC 668
Db 529 TGCCCCAAATGCTTGTGTCACCATCATGTGCTGGCATGAAGGATCTGTGATCCACGC 588
QY 669 CATGAAGTGGGCCAGGACCACTGTGAGCGCTCGCAAGAACAGCTTTGAGCTCTACGGGGC 728
Db 589 ACTTCAGACCTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTATGGCGC 648
QY 729 TGACTTGTCTTGGAGGGAATTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCCAC 788
Db 649 TGACTTGTGTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGCCCCAC 708
QY 789 CATGACCCGTCACCGCGTCAACGCCCGCTGTGTCACAGGTGTCAGAGGACGACACCAT 848
Db 709 GATGGCACCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGGTGCAAGCTGACACCT 768
QY 849 CAAAGTGGCGGTGGACCGCA 868
Db 769 GCGGTGGTTCATGACCGGA 788

RESULT 9
AK080321
LOCUS
DEFINITION
AK080321 2520 bp mRNA linear HTC 03-APR-2004
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
library, clone:A630053H17 product:HOTTL PROTEIN homolog [Homo
sapiens], full insert sequence.
AK080321
AK080321.1 GI:26348476
HTC; CAP trapper.
MUS
Mus musculus (house mouse)
MUS
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
REFERENCE
AUTHORS
Carninci, P., Shibata, Y., Hayateu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
1042159
3
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimura, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer

JOURNAL MEDLINE PUBMED REFERENCES AUTHORS	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861 4	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection	ORIGIN	Query Match Best Local Similarity 41.6%; Score 377; DB 3; Length 2520; Matches 566; Conservative 0; Pred. No. 4.1e-80; Mismatches 315; Indels 0; Gaps 0;	VIVPMKAAVIAHALQTSQDNVOCRKASFELYGADVFVGFEDFOPWLEINASPTMAPST AVTARLCAGVQADTLRLVDRDLDRSCDTGAFELIYKQPAVEVPPVQVIGILLVEGSTI KKFVPVGHRTGVRSSLSPLHLTQGGSGESKDSGSPHRSASRKNARAESLEHTEKPEP AAVASVGKGAAPFHPFSLHSAWLPSPRVHRPQGRVLRQLQDLQDLVGSKALSTTGKA LMTLPTAKVLMSPFHPDLKLAFLPSMLKPGKVGFELOCCTTWRVVLSSGIGSEGHQRAA PRSSAPGKGLSSTEPSKSTET"
JOURNAL REFERENCE AUTHORS	Nature 409, 685-690 (2001) 5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	QY	2 AGGACATCGACAGCTCAGCAGATCGCGTGAGGACCTCAGTGAGCGCGAGTGGGAGGACC 61 DB	631 AAGACATTCAGCAGGACCCCAACTCCCGCTCTACCTCAGCCCGATGATGGTCCAGT 690
JOURNAL REFERENCE AUTHORS	Nature 420, 563-573 (2002) 6 (bases 1 to 2520)	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.	QY	62 TGACCCAGCAGTACTACTCCCTCGTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAAT 121 DB	691 TCCTTCAGCGCTACTACCAATAGTTTCATGAAGGGGAGAGCTCAGATACCTCGAGGTCC 750
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	122 ACTTTTCGAGTCCAGGCTCTCTGTAATAGATCAGCTCTGTGAACCTTCAGACGACA 181 DB	751 AAGTCCAGCGCTGTGAAGACATCTCAGCAGCTTCAGAACTGGTACCCAGTGTAGACA 810
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	182 TTGACGGGCTCCGGAACATCTGGATTATAAGCCCGCGGCAAGTCCCGGGCGGAGACA 241 DB	811 TGGAGGGGATCGGAACATCTGGATCGTGAAGCCCGGAGCCAGTCCCGAGGCGGAGGA 870
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	242 TAGTGTGCATGAGACCGTGTGGAGGAGATCTGGAGCTGGGAGTGCAGACCAACCTCTTT 301 DB	871 TTATGTGCATGAACCGCTGATGATGATGCTGAAGCTGGTGGAGTGCACACCCCATGCTCA 930
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	302 CGAGGGAACAAGTGGTGGTTCAGAACTACATCAGACGCGCTGCTCATCTGTGACA 361 DB	931 TGAAGGATGCAAGTGGATCGTCAGAACTGATTCAGCGGCGCTCTTATCTTCGGCA 990
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	362 CCAAGTTTCAGCATCAGACAGTGGTCTCTCTCAGGAGTGGAAACCCCTGACCACTCTGGT 421 DB	1051 TCTACCCGAGACGACGATCATTTCTCTTCTCAGACAGCCCTTCTCCTCGAAGAACCTGGACA 1110
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	482 GCGCCATCAGCTGTGCAACAACGCGTTCAGAACTACCTGGAAGATGATGTGGGCGCGCA 541 DB	1111 ACTCTGTGACCTGTGTAACTCCATCCATCCAGAGACACTTGGAGGCTTCTCTGTCACCGGC 1170
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	542 GCGCCCTGCTGCGCGCAGACACATGTGGACAGCAGCAGCTTCCAGGAGTACCTGCGAGC 601 DB	1171 ACCGATGCTGCGCGCAGATTAACATGTGGTCCAGCAGAGGTTTCAGGCCCATCTTGAGG 1230
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	602 GCCAGGCGCTGCGCGCTGTGGGCGAGCTCATCTACCCGTCATGGAAGAGGCCATCG 661 DB	1231 AGGTAGATGCGCGCAAGGCTGCTTCAGCGTTCATCGGCCAGGATGAAGGCTGCTGTA 1290
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	662 CCCAGCATGAGAGTGGCGCCAGGACCACTGAGCGCTTCGCAAGAACAGCTTTGAGCTCT 721 DB	1291 TCCATGCGCTGCGAGCCTCCCAAGATAACGTGAGCGTGCAGGAGGCCAGCTTTGAGCTCT 1350
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	722 ACGGGCTGACTTCTGCTTGGGAGGAGCTTTCAGGCGCTGCTGATCGAGTCAATTCCTCA 781 DB	1351 ATGGGCGAGACTTTGTTGTTGGGGAAGACTTTCAGGCGCTGTTGATTTGAATCAATGCCA 1410
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	782 GCCCCACCATGACCCGCTCCACCGCTCAGGCGCCAGCTGTGTGACAGGTGCGAGGAG 841 DB	1411 GCCCCACCATGCGACCTTCCACGGCTGCTCACTGCCCGCTCTGTGCCGCTGTGCGAGG 1470
JOURNAL REFERENCE AUTHORS	Submitted (16-APR-2002)	Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	QY	842 ACACCATCAAGTGGCGCTGGAACCGCAGCTGTGTGACATCGCG 882 DB	1471 ATACCTGCGTGTGGTTCATCGACCGCGCTGCGACCGCAGC 1511

Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)
5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)
6 (bases 1 to 2520)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

Submitted (16-APR-2002)
Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

Location/Qualifiers

1. 2520
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:A630053H17"
/db_xref="taxon:10090"
/clone="A630053H17"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
84. 2198
/note="unnamed protein product; HOTTL PROTEIN homolog [Homo sapiens] [SPTR]Q9UI99, evidence: FASTA, 85.8%ID, 93.4%length, match=821)
putative"
/codon_start=1
/protein_id="BAC37878.1"
/db_xref="GI:26348477"

TRANSLATION="MDTSSPPYAGQGEASPPSLPVSISSPSSQSRVNRNETPYLIWT
AEDDKAFIDFVMTAARNVLKLVKLEKSQSIQAREEAPEDTPKKQKGLVT
VSSDFVDEALSAQOHLSSIAHKIDIKDPNSPLYLSDDWSQFLQYQIVHSGAELR
YLVEQVORCDILQOLQNVLPOLMEGRNIWIKPGAKSRGRGIMCNRDLBMLKLV
DCHPLMKDGMVQKVIETPLIFGTFKDLROWFLVTDNPLVTWVYRDSYIRFSTQ
PFSLKLNLDNSVHLNCNLSIQHLEASCHRHPLMLPPDNMWSQRFOAHLQEVDAKAWSS


```
RESULT 10
BU937372
LOCUS
DEFINITION
AGENCY 10519180 NIH_MGC_169 Mus musculus cDNA clone
IMAGE:6705256 5', mRNA sequence.
BU937372
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
CDNA Library Preparation: Michael Brownstein Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM3002 row: 9 column: 16
High quality sequence stop: 569.
FEATURES
source
1..775
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6705256"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_169"
/note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI
(ggccattggcc); Site 2: SfiI (ggcgctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCATGGTATCAACGAGTGGCCATTACGCGGG-3' and
5'-ATTCTAGAGCCGAGCGGCGGACATG-dT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
ORIGIN
Query Match 40.8%; Score 370; DB 5; Length 775;
Best Local Similarity 81.2%; Pred. No. 1.7e-78;
Matches 446; Conservative 0; Mismatches 91; Indels 12; Gaps 1;
QY 370 GACATCAGACAGTGGTTCTCTGTCACGAGTGAACCCCTGACCATCTGGTTCTACAAG 429
DB 1 GGCATCAGACAGTGGTTCTCTGTCACAGATGGAATCCCTACCATCTGGTTCTACAA 60
QY 430 GAGAGTTACTTGGCGTTCTCAACTCAGCGGTTCTCCCTCGACAGAGTGGACAGCGCATC 489
DB 61 GAAAGCTACTTCGCTTCTCCACACAGCGCTTCTCCCTCGACAGCTGGACAGCGCATC 120
QY 490 CACCTGTGMAACACCGCGTTCGAGATGACTGAAGATGATGGCGCGCAGCCCGCTG 549
DB 121 CACCTGTGTAACAACATCCATCCAGAGCGCGCTCAAAAATGATAAAGAGCGAGTCGCGTG 180
QY 550 CTGCCCGCACACAACATGTGGACACACACAGGTTCCAGAGTACCTGCGACGCCAGGGC 609
DB 181 CTACCTGTGCATACATGTGGACACAGCATCTGTTTCCAGAGTACTTCGAGAAGAGGGC 240
QY 610 CGTGGCGCGGTGTGGGCGACGCTCATCTACCCGTCCATGAAGAGCCATCGCCACGCC 669
DB 241 CGAGGAGGAACGTGGGGTAGCATCATCTACCCGTCTATGAAAAGAGCTGTCAACCAATGC 300
QY 670 ATGAAGGTGCCCGACGACCAACGTGGAGCCCTCGCAAGACAGCTTTCAGCTCTACGGGCT 729
|||||
```

```
DB 301 ATGAGGGTAGCCAGGACCAACGTAAGCCCGTAAGACAGAGCTTCGAGCTCTACGGAGCC 360
QY 730 GACTTCGTCTCTGGAGGAGCACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCGCCACC 789
DB 361 GACTTCATCTCTGGGCGAGACTTCAGCCATGGCTTATCGAGATCACTTCAGCCGCCACC 420
QY 790 ATGACACCCCTCCACGCGGTTCAGCGCCAGCTGTGTGTCACAGGTGCGAGGAGCACCATC 849
DB 421 ATGACACCCCTCCACCTCCGTCACAGCTCTGTGTCAGGTGCGAGGAGGACCATC 480
QY 850 AAGTGGCGCG-----TGGACCCAGCTGTGACATCGGCAACTTCGAGCTCTCG 897
DB 481 AAGTGGTGGTGGATCGCAAACTGGACCGAAATGTGACATCGGCAACTTTGAGCTTCTG 540
QY 898 TGGAGGCGAG 906
DB 541 TGGCGGCGAG 549
RESULT 11
LOCUS
DEFINITION
AGENCY 6617783 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734160
IMAGE:5734160
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM12739 row: a column: 09
High quality sequence start: 78
High quality sequence stop: 658.
FEATURES
source
1..1022
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734160"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(ddestroyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."
ORIGIN
Query Match 39.7%; Score 360; DB 4; Length 1022;
Best Local Similarity 68.8%; Pred. No. 4.6e-76;
Matches 495; Conservative 0; Mismatches 225; Indels 0; Gaps 0;
QY 129 GCAGTGGCCAGCTCTCTGAATAGATACAGTCTGTGAACCTTCAGACGACATTACCG 188
DB 40 GCGCTGTGAGGACATCTCTCAGACAGCTGCAGCGCGGTGATCCCATAGATAGACGAGG 99
|||||
```



```
Qy 189 GCTCCGGACATCTGGATTAAAGCCGCGGCAAGTCCCGGGCGGAGACATAGTGTG 248
Db 100 GGATCGCAACATCTGGATCGTGAAGCAGGAGCAAGTCCCGGAGGAGCATCATGTG 159
Qy 249 CATGGACCGTGTGGAGGAGATCTCGAGCTGGCAGCTGCAGACCACTCTTTCACAGGA 308
Db 160 CATGGACCACTCGAGGAGATCTGAAGCTGTGAACGGCAACCCCGTGGTGAAGGA 219
Qy 309 CAACAAGTGGTGTGCAGAAAGTATCATCGAGACCGCGTGTCTCATCTGTGACACCAAGTT 368
Db 220 CGGCAAGTGGTGTGCAGAAAGTATCATCGAGACCGCGTGTCTCATCTTGGCACCAGTT 279
Qy 369 CGACATCAGACAGTGGTTCCTCGTCAGGACTGGAACCCCTGACCATCTGTTCTACAA 428
Db 280 TGACCTCTTACAGTGGTTCCTCGTGAACCTGTAACCTTACCGTGGTGTCTACCG 339
Qy 429 GGAGAGTACTTGGCGGTCTCAACTCAGCGCTTCTCCCTGACAAAGTGGACAGCGCAT 488
Db 340 CGACAGTATATCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAACCTGGACAACTCAGT 399
Qy 489 CCACCTGTGCAACAAACCGGTTCAGAAAGTACTCTGAAGAAATGATGGGCGCAGCCCT 548
Db 400 GCACCTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCTCCATGCCACT 459
Qy 549 GCTCCCGGCACACACATGTGGACAGCAGCCAGTTCGAGAGTACTCGAGCGCCAGGG 608
Db 460 GTTTCGCCAGACAAATGTGGTCTAGCCAGAGTTCAGGCCACCTGCGAGGATGGG 519
Qy 609 CCGTGGCGCGTGTGGGCGAGCGTCACTACCGCTCATGAAGAGCCATCGGCCACGC 668
Db 520 TGCCCCAAATGTTGGTCCACCATCATCTGTCGCTGGCATGAGATGCTGTGATCCAGC 579
Qy 669 CATGAAGTGGCCAGACCACTGGAGCTTCGCAAGAACAGTTTGAAGTCTACGGGG 728
Db 580 ACTTCAGACCTCCAGGACACCGTGCAGTGTGGAAGCCAGCTTTGAGCTCTATGGCG 639
Qy 729 TGACTTCGTCTTGGAGGAGTTCAGGCCCTGCTGATCGAGATCAATTCAGCCCCAC 788
Db 640 TGACTTCGTGTGGGAGGAGTTCAGGCCCTGCTGATTCAGATCAACCGCCGCCAC 699
Qy 789 CATGACCCCGTCCACCGCGGTTCAGGCCAGCTGTGTGCACAGTGCAGGAGGACACCAT 848
Db 700 GATGGCACCCCTCACAGAGTCACTGGCGCGCTCTGTGCTGGGTGCAAGCTGACACCT 759
```

```
RESULT 12
BM717061
LOCUS BM717061 693 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-EJ0-ahk-e-05-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone
BM717061
VERSION BM717061
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
1 (bases 1 to 693)
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
```

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1..693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-ahk-e-05-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJ0"
/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an Ecor I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes,
AGATCAACA; lens, CGATTAGCGA; eye anterior segment,
AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG;
Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
library was created for the program, Gene Discovery in the
Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 35.4%; Score 320.6; DB 4; Length 693;
Best Local Similarity 69.1%; Pred. No. 1.4e-66;
Matches 472; Conservative 0; Mismatches 199; Indels 12; Gaps 2;
Qy 235 CGAGACATAGTGTGACCGTGTGGAGGAGATCTCGAGCTGGCAGCTGCAGACCA 294
Db 1 CGAGGCATCATGTGTCATGACCACTGGAGGAGATGCTGAAGCTGGT-GAACGCA 59
Qy 295 CCTCTTTCAGGACCAACAAGTGGTGGTCCAGAGTACATCGAGACCGCTGCTCATC 354
Db 60 GTGGTGAATGAAGACGCGCAAGTGGTGGTGCAGAGTATATTGAGCGGCCCTCTCATC 119
Qy 355 TGTGACACCAAGTTCGACATCAGACAGTGGTTCCTGTCAGGACCTGGACCCCTGACC 414
Db 120 TTTGGCACCAGTTTGAACCTCAGACAGTGGTTCCTGGTAACTGACTGGAAACCACTTACC 179
Qy 415 ATCTGGTTTCAAGGAGAGTACTTTCGGTTCCTCAACTCAGGCTTCTCCCTGGACAAG 474
Db 180 GTGTGTTTACCGGACAGCATATATCCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAAC 239
Qy 475 CTGGACAGCCCATCCACCTGTGCAACAACCGCTCCAGAAAGTACCTGGAAGATGATGTG 534
Db 240 CTGGACAACCTCAGTGCACCTGTGCAACAACCTCCATCCAGAAGCACCTGGAGATCATGC 299
Qy 535 GGGCGGAGCCCTGCTGTCGCCGACACAAATGTGGACACAGCACAGGTTCCAGGAGTAC 594
Db 300 CATCGGCATCCACTGCTTCGCCAGACAAATGTGGTCTAGCCAGAGGTTCCAGGCCAC 359
Qy 595 CTCGAGCGCCAGGCGCGTGGCGCGGTGTGGGAGGCTCATCTACCGCTCCATGAAGAAG 654
Db 360 CTCGAGAGATGGTGGCCCAATGTGTTGTCACCAATCATCTGTCCTGGCATGAAGAT 419
Qy 655 GCCATGCCCGCCCATGAGAGTGGCCAGGACCACTGGAGGCTTCGAGAAACAGCTTT 714
Db 420 GCTGTGATCCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGCGAAGGCCAGCTTT 479

QY 715 GAGCTCTACGGGGCTGACTTCCTTGCTTGGAGGGAGCTTCAGGCCCTGGCTGATCGAGATC 774
 Db 480 GAGCTCTATGGCGCTGACTTCGTTGTTGGGGAGGAGCTTCAGGCCCTGGCTGATGAGATC 539
 QY 775 AATTCAGCCCCCAGCAGCAGCGTCCAGCCCGGTGACGCCCGGAGCTGTGTGACAGGTG 834
 Db 540 AAGCCAGCCCCCAGCAGTGGCAGCCCTCCACAGCAGTCACTGCCCGGCTCTGTGTGGCGTG 599
 QY 835 CAGGAGGACAC-----CATCAAGGTGCGCGTGGAGCCGAGCTGTGACATCGGCA 883
 Db 600 CAGCTGAACCTGCGCGGTGCTATTGACCGGATGCTGGACCGCAACTGTGACACAGGAG 659
 QY 884 ACTTCGAGCTCTGTGGAGGACG 906
 Db 660 CTTTGAGCTCATCTATAAGCAG 682

RESULT 13
 CK653583 861 bp mRNA linear EST 29-JAN-2004
 AGENCOURT 17673480 NIH MGC 237 Rattus norvegicus cDNA clone
 IMAGE:7113394 5', mRNA sequence.

CK653583
 CK653583.1 GI:41389106
 EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 861)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LLNL4985 row: e column: 08

High quality sequence stop: 304.

Location/Qualifiers

1. .861

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="IMAGE:7113394"

/issue_type="testis, pooled"

/lab_host="DH10B Tona"

/clone_lib="NIH MGC 237"

/note="Organ: testis; Vector: pExpress-1; Site 1: EcorV;

Site_2: Not; RNA obtained from testis tissue of 8 wk old animal. Tissues were snap-frozen and kept at -80C before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer:

5'-pGACTAGTTCTAGTCGAGCGGCCCGCC(T)25-3' and cloned into the EcorV/NorI sites of pExpress-1. Size-selection 1.4kb resulted in an average insert size of 2.4 kb. This primary library is not normalized (normalized primary library is NIH MGC 238) and was constructed by Express Genomics (Frederick, MD)"

ORIGIN

Query Match 34.9%; Score 316; DB 7; Length 861;

Best Local Similarity 74.2%; Pred. No. 1.9e-65;

Matches 439; Conservative 0; Mismatches 150; Indels 3; Gaps 3;

QY 227 CCGGGGCGGAGACATAGTGTGCATGGACCGTGTGAGGAGATCCTTGGAGCTGGCAGCTG 286
 Db 12 CCGGGATTGGAGATATTGTGTCATGGACCGTGTGAGAACATCCTTGGATCTGTGGCC- 70
 QY 287 CAGACCAACCTCTTTTCCAGGGACAACAAGTGGGTGTGCAGAAAGTACATCGAGACGCCGC 346
 Db 71 CAGACAGGAGAGACCAAGAGGACAACAATGGGTGTGCATTAAGTACATCGAGACCCCGA 130
 QY 347 TGTCTATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCTCTGTCAGGACTCGGAACC 406
 Db 131 TGTCTATCTATGACCAAGTTTGACATCAGACAGTGGTTCTTGTACAGACTGGAATC 190
 QY 407 CCTGACCATCTGGTTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCC 466
 Db 191 CCTTAACCATCTGGTTCTACAAGAGAGCTACCTCCGTTCTCCAAGCGCTTCTCCC 250
 QY 467 TGGACAAAGTGGACAGCGGCATCCACTGTGCAACAACCGCGTCCAGAGTAGTCTGAAGA 526
 Db 251 TGGACAAACTGGACAGCGCTTCCACCTGTGTAACTCCGTCCAGAGCGTGGCATAA 310
 QY 527 ATGATGTGGGCGCAGCCCTCTGCTGCCCGCACACAACATGTGGACCGACAGCAGGTCC 586
 Db 311 ATGACAAAAACGAAGTCCACTGCTACCTTGCCTAACATGTGGACCATCACCCTTCC 370
 QY 587 AGGAGTACCTGCAGCGCCAGGGCCGTGGCGCGCGTGTGGGCGCAGCGTCATCTACCCGTCCA 646
 Db 371 ATGAGTACTCTGANTATAGGGCCGAGGATGACATGGGGTAGTATCACTACCCGCTCA 430
 QY 647 TGAAGAAGGCCATCGC-CCACGCCATGAAGGTGGCCAGGACCAACGTCGAGCGCTCGCAAG 705
 Db 431 TGAAGAAGAGCTGTCACCTCAACGTCATGATGGTGGCCCAAGATCTTGTGAAGCCCGTAAG 490
 QY 706 AACAGCTTTGAGCTCTACGGGCTGACTTCGTCTCTGGGAGGACTT-CAGGCGCTGGCT 764
 Db 491 ATTATCTTCGATCTCTATGGAGCTGACTTTAATCTGGGCGTGACTTATAGTCTCTGTGT 550
 QY 765 GATCGAGATCAATTCAGGCCCCACCATGCACCGCGTCCAGCGCGTCCACGCGCC 816
 Db 551 CTTCAAGATCAACTTCTACTCTGCTTGCACCTCTACTTTGTAAAMACC 602

CD515431 896 bp mRNA linear EST 06-JUN-2003
 AGENCOURT 14364765 NIH MGC 181 Homo sapiens cDNA clone
 IMAGE:30407505 5', mRNA sequence.

CD515431
 CD515431.1 GI:31447149
 EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 896)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Michael Brownstein

CDNA Library Preparation: Invitrogen Corp

DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: NDM488 row: p column: 10

High quality sequence stop: 678.

Location/Qualifiers

FEATURES

RESULT 15	CK597795	777 bp	linear	EST 22-JAN-2004
LOCUS	AGENCOURT_17837841	NIH MGC 238		
DEFINITION	IMAGE:7135519.5', mRNA sequence.			
ACCESSION	CK597795			
VERSION	CK597795.1	GI:41110910		
KEYWORDS	EST.			
SOURCE	Rattus norvegicus	(Norway rat)		
ORGANISM	Rattus norvegicus			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			

Qy 421 TTCTACAAGGAGAGTTACTTGGCGTTCTC 449
| | | | | | | | | | | | | | | | | |
Db 671 TTCTACAAGGAGAGCTACCTCGGTTCTC 699

Search completed: September 25, 2005, 01:56:03
Job time : 3483.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 01:51:04 ; Search time 549.716 Seconds
(without alignment)
9756.464 Million cell updates/sec

Title: US-10-615-659-1_COPY_369_1274

Perfect score: 906

Sequence: 1 gaggacatcgacgtcagc.....tcgagctctctgtggagcag 906

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N Geneseq_16Dec04:*
2: Geneseqn1980s:*
3: Geneseqn1990s:*
4: Geneseqn2000s:*
5: Geneseqn2001as:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	906	100.0	1838	ADJ93357	Adj93357 Human BGS
2	906	100.0	3554	ADJ93364	Adj93364 Human tub
3	806.8	89.1	1859	ADJ93362	Adj93362 Human BGS
4	762	84.1	3465	ADJ93363	Adj93363 Human BGS
5	760.4	83.9	1939	ADJ93361	Adj93361 Human BGS
6	726	80.1	726	ADJ93359	Adj93359 Human BGS
7	427	47.1	490	ACH91699	Ach91699 Human gen
8	427	47.1	101270	ADQ17814	Adq17814 human sof
9	400.4	44.2	2881	AAC77214	Aac77214 Human ORF
10	398.8	44.0	2326	AAH16735	Aah16735 Human cDN
11	398.8	44.0	2380	AAS99894	Aas99894 Polynucle
12	398.8	44.0	2538	ADS73168	Ads73168 Human kid
13	398.8	44.0	2553	ACN37881	Acn37881 Tumour-as
14	398.8	44.0	2848	ADM03081	Adm03081 Human cDN
15	398.8	44.0	3001	AAI58606	Aai58606 Human pol
16	398.8	44.0	3001	ADQ98824	Adq98824 DNA encod
17	398.8	44.0	3001	ADB48584	Adb48584 Novel hum
18	398.8	44.0	3294	ABX34502	Abx34502 Human mdd
19	390.6	43.1	2611	AAF21812	Aaf21812 Human bre
20	329.8	36.4	2412	ADH45468	Adh45468 Human mol

21	243.2	26.8	2469	12	ADH45478	Adh45478 Human mol
22	212.2	23.4	5282	12	ADQ64426	Adq64426 Novel hum
23	211.8	23.4	2044	4	AAI60392	Aai60392 Human pol
c	208.6	23.0	3828	13	ADR07582	Adr07582 Full leng
25	207.4	22.9	2979	4	ABL17657	Ab117657 Drosophil
26	190.6	21.0	1958	4	AAH18689	Aah18689 Human cDN
c	190.6	21.0	10468	4	AAK77961	Aak77961 Human imm
28	189.2	20.9	2217	10	ACA92443	ACA92443 DNA encod
29	189	20.9	1567	8	ACC46592	ACC46592 Human dit
30	174.2	19.2	2543	4	ABL17449	Ab117449 Drosophil
31	166	18.3	1728	5	AAS70158	Aas70158 DNA encod
32	164.4	18.1	2250	5	AAS69383	Aas69383 DNA encod
33	163.4	18.0	418	8	ABX52103	Abx52103 Bovine ES
34	144.8	16.0	4615	4	ABL17448	Ab117448 Drosophil
c	136.8	15.1	5748	4	ABL17656	Ab117656 Drosophil
36	131.8	14.5	1085	4	AAF63819	Aaf63819 Human sec
37	130	14.3	525	12	ACH77979	Ach77979 Human gen
38	117	12.9	492	9	ACH48041	Ach48041 Human lun
39	92.2	10.2	207	6	ABS69387	Abs69387 Novel mur
c	80	8.8	80	12	ADJ933402	Adj933402 Human BGS
41	79.4	8.8	1571	4	ABL14515	Ab114515 Drosophil
42	71.8	7.9	996	10	ADC32299	Adc32299 Human nov
43	71.8	7.9	2259	12	ADQ64293	Adq64293 Novel hum
44	71.8	7.9	2417	10	ADC30440	Adc30440 Human nov
45	65.6	7.2	1579	8	ABX05228	Abx05228 Human nov

ALIGNMENTS

RESULT 1
ADJ93357
ID ADJ93357 standard; cDNA; 1838 BP.

XX AC ADJ93357;

DT 06-MAY-2004 (first entry)

XX DE Human BGS-42 cDNA sequence SeqID1.

XX KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; neurotropic; antiparkinsonian; anticholinergic; antidiabetic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine kinase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.

XX KW Homo sapiens.

Key Location/Qualifiers
CDS 153..1778
FT /*tag= a
FT /product= "Human BGS-42 protein"

XX WO2004005487-A2.

XX 15-JAN-2004.

PF 09-JUL-2003; 2003WO-US021605.

PR 09-JUL-2002; 2002US-0394725P.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX MPI; 2004-099381/10.
DR P-PSDB; ADJ93358.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
PS
XX
XX Claim 1; SEQ ID NO 1; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.
XX
SQ Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;
Query Match 100.0%; Score 906; DB 12; Length 1838;
Best Local Similarity 100.0%; Pred. No. 2.1e-207;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGGACATCGACACGCTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAC 60
DB 369 GAGGACATCGACACGCTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGAGGAC 428
QY 61 CTGACCCAGCAGTACTACTCTCTGTTTCATGGGATGCTTTTCACTCCAAATTCAGAAAT 120
DB 429 CTGACCCAGCAGTACTACTCTCTGTTTCATGGGATGCTTTTCACTCCAAATTCAGAAAT 488
QY 121 TACTTTTCCAGTGGCCAGGCTCTGCTGATAGAAATCAGCTCTGTGAACCTCAGACGGAC 180
DB 489 TACTTTTCCAGTGGCCAGGCTCTGCTGATAGAAATCAGCTCTGTGAACCTCAGACGGAC 548
QY 181 ATTGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCCGCAAGTCCCGGGCCGAGAC 240
DB 549 ATTGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCCGCAAGTCCCGGGCCGAGAC 608
QY 241 ATAGTGTGCATGCACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT 300
DB 609 ATAGTGTGCATGCACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT 668
QY 301 TCCAGGGACAACAGTGGGTGTCAGAGATACATCGACGCCCGTCTCATCTGTGAC 360
DB 669 TCCAGGGACAACAGTGGGTGTCAGAGATACATCGACGCCCGTCTCATCTGTGAC 728
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGGAAACCCCTGACCACTCTGG 420
DB 729 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACCGACTGGAAACCCCTGACCACTCTGG 788
QY 421 TTCTACAAGGAGAGTTTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 480

DB 789 TTCTACAAGGAGAGTTTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 848
QY 481 AGCGGCATCCACTCTGTGCAACCAACCGCGTCCAGAAAGTACTCTGAAGAAATGATGTGGGCCCG 540
DB 849 AGCGGCATCCACTCTGTGCAACCAACCGCGTCCAGAAAGTACTCTGAAGAAATGATGTGGGCCCG 908
QY 541 AGCCCCCTGCTGCCCGCACACAACATGTGGACACAGCAGGTTCAGGAGTACTCTGCAG 600
DB 909 AGCCCCCTGCTGCCCGCACACAACATGTGGACACAGCAGGTTCAGGAGTACTCTGCAG 968
QY 601 CGCCAGGGCCGTTGGCGCGTGTGGGGCAGCGTCACTACCCGTCATGAGAGGCCATC 660
DB 969 CGCCAGGGCCGTTGGCGCGTGTGGGGCAGCGTCACTACCCGTCATGAGAGGCCATC 1028
QY 661 GCCCAGCGCCATCAAGGTGGCCCGCAGACCAACGCTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 720
DB 1029 GCCCAGCGCCATCAAGGTGGCCCGCAGACCAACGCTGGAGCCCTCGCAAGAACAGCTTTGAGCTC 1088
QY 721 TACGGGGCTGACTTCTGCTCTTGGGAGGAGATTTCAGGCCCTGCTGATCGAGATCAATTCC 780
DB 1089 TACGGGGCTGACTTCTGCTCTTGGGAGGAGATTTCAGGCCCTGCTGATCGAGATCAATTCC 1148
QY 781 AGCCCCCATCATCCCGTCCAGCGCGTCCAGCGCCAGCTGTGTCACAGTGTGCAGGAC 840
DB 1149 AGCCCCCATCATCCCGTCCAGCGCGTCCAGCGCCAGCTGTGTCACAGTGTGCAGGAC 1208
QY 841 GACACCATCAAGGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 900
DB 1209 GACACCATCAAGGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 1268
QY 901 AGGCAG 906
DB 1269 AGGCAG 1274
RESULT 2
ADJ93364
ID ADJ93364 standard; DNA; 3554 BP.
AC ADJ93364;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.
OS Homo sapiens.
OS Synthetic.
XX
FH Location/Qualifiers
FT 1695..3320
FT /*tag= a
FT /product= "Human tubulin tyrosine ligase protein"
XX
XX WO2004005487-A2.
XX 15-JAN-2004.
XX

PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
PI
XX WPI; 2004-099381/10.
DR P-PSDB; ADJ93365.
DR
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Example 4; SEQ ID NO 12; 343pp; English.
PS
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytosolic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of the
CC tubulin tyrosine ligase protein consensus gene sequence which was used in
CC the exemplification of the invention.
XX
SQ Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;
Query Match 100.0%; Score 906; DB 12; Length 3554;
Best Local Similarity 100.0%; Pred. No. 2.5e-207;
Matches 906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGGACATCGACAGCTCAGCAGATGCCGTGGAGGACCTCACTAGGCCGAGTGGAGGAC 60
DB 1911 GAGGACATCGACAGCTCAGCAGATGCCGTGGAGGACCTCACTAGGCCGAGTGGAGGAC 1970
QY 61 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATTCCTCAATCAAGAAAT 120
DB 1971 CTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATTCCTCAATTCAGAAAT 2030
QY 121 TACTTTTCGAGTGCAGGCTGCTGCTGAATAGATCACTGCTGACCTCTCAGACGGAC 180
DB 2031 TACTTTTCGAGTGCAGGCTGCTGCTGAATAGATCACTGCTGACCTCTCAGACGGAC 2090
QY 181 ATTGACGGGCTCCGGAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 240
DB 2091 ATTGACGGGCTCCGGAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGGCCGAGAC 2150
QY 241 ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCTCTT 300
DB 2151 ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCTCTT 2210
QY 301 TCCAGGAGCAACAAGTGGGTGGTCCAGAAATACATCGAGAGCGCGGTCTCTCTGTGAC 360

DB 2211 TCCAGGAGCAACAAGTGGGTGGTCCAGAAATACATCGAGAGCGCGGTCTCATCTGTGAC 2270
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTTGACCATCTGG 420
DB 2271 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTTGACCATCTGG 2330
QY 421 TTCTACAAGGAGAGTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGAC 480
DB 2331 TTCTACAAGGAGAGTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGAC 2390
QY 481 AGCGCCATCCACCTGTGCAACCAACGCGCTCCAGAGTACCTGAAGAATGATGTGGGCGGC 540
DB 2391 AGCGCCATCCACCTGTGCAACCAACGCGCTCCAGAGTACCTGAAGAATGATGTGGGCGGC 2450
QY 541 AGCGCCCTGTGCTGCCGCAACAACATGTGGAACAGCAGCAACAGGTTCCAGGAGTACCTGCAG 600
DB 2451 AGCGCCCTGTGCTGCCGCAACAACATGTGGAACAGCAGCAACAGGTTCCAGGAGTACCTGCAG 2510
QY 601 CGCCAGGCGGTGGCGCGCTGTGGGCGAGGTCATCTACCGTCCATGAGAGAGGCCATC 660
DB 2511 CGCCAGGCGGTGGCGCGCTGTGGGCGAGGTCATCTACCGTCCATGAGAGAGGCCATC 2570
QY 661 GCCCAGCCATGAAGTGGCCAGAGCACAGTGGAGCCTCGCAAGAACACAGCTTTGAGCTC 720
DB 2571 GCCCAGCCATGAAGTGGCCAGAGCACAGTGGAGCCTCGCAAGAACACAGCTTTGAGCTC 2630
QY 721 TAGGGGCTGACTTCTCTTGGAGGGACTTCAGGCCCTGGCTGATTCGAGATCAATTC 780
DB 2631 TAGGGGCTGACTTCTCTTGGAGGGACTTCAGGCCCTGGCTGATTCGAGATCAATTC 2690
QY 781 AGCCCCCACCATGACCCGCTCCAGCGGTCACGGCCAGCTGTGTCACAGGTGCGAGGAG 840
DB 2691 AGCCCCCACCATGACCCGCTCCAGCGGTCACGGCCAGCTGTGTCACAGGTGCGAGGAG 2750
QY 841 GACACCATCAAGTGGCCGCTGGAGCCAGCTGTGACATCGGCAACTTCAGAGCTCTCTGG 900
DB 2751 GACACCATCAAGTGGCCGCTGGAGCCAGCTGTGACATCGGCAACTTCAGAGCTCTCTGG 2810
QY 901 AGGCAG 906
DB 2811 AGGCAG 2816
RESULT 3
ADJ93362
ID ADJ93362 standard; DNA; 1859 BP.
XX
AC ADJ93362;
XX
DT 06-MAY-2004 (first entry)
XX
XX Human BGS-42 protein-related DNA clone B SeqID10.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiaesthetic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX
OS Homo sapiens.
XX
XX WO2004005487-A2.
PN

PD 15-JAN-2004.
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
PS Example 4; SEQ ID NO 10; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, neurotropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX
SQ Sequence 1859 BP; 418 A; 563 C; 562 G; 316 T; 0 U; 0 Other;

Query Match 89.1%; Score 806.8; DB 12; Length 1859;
Best Local Similarity 95.0%; Pred. NO. 1.3e-183;
Matches 861; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

QY 1 GAGGACATGCAGACGTCAGCAGATCCGCTGGAGGACCTCACTGAGCCGAGTGGAGGAC 60
DB 263 GAGGACATGCAGACGTCAGCAGATCCGCTGGAGGACCTCACTGAGCCGAGTGGAGGAC 322
QY 61 CTGACCCAGCAGTACTACTCCCTGTTCTATGTCGCGATGTTTCACTCTCCAATTCAGAAAT 120
DB 323 CTGACCCAGCAGTACTACTCCCTGTTCA----- 351
QY 121 TACTTTTCGAGTGCAGGCTCTGTGAATAGATCACTCTGTGAACCTCAGACGGAC 180
DB 352 -----TGCCAGGCTCTGCTGAATAGATCACTCTGTGAACCTCAGACGGAC 399
QY 181 ATTGACGGCTCCGGAACATCTGGATTATAAGCCCGCGGCAAGTCCCGGGCCGAGAC 240
DB 400 ATTGACGGCTCCGGAACATCTGGATTATAAGCCCGCGGCAAGTCCCGGGCCGAGAC 459
QY 241 ATAGTGTGATGACCGCTGTGGAGAGATCTCTGGAGCTGGCAGCTCAGACCAACCTCTT 300
DB 460 ATAGTGTGATGACCGCTGTGGAGAGATCTCTGGAGCTGGCAGCTCAGACCAACCTCTT 519
QY 301 TCCAGGGACAACAAAGTGGTGGTCCAGAAAGTACATCGAGACGCGCGCTCTCATCTGTGAC 360

DB 520 TCCAGGGACAACAAAGTGGTGGTCCAGAAAGTACATCGAGACCGCGCTCTCATCTGTGAC 579
QY 361 ACCAAAGTTTCGACATCAGACAGTGGTTCCTCGTCA CGGACTCGGAACCCCTGACCACTCTGG 420
DB 580 ACCAAAGTTTCGACATCAGACAGTGGTTCCTCGTCA CGGACTCGGAACCCCTGACCACTCTGG 639
QY 421 TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 480
DB 640 TTCTACAAGGAGAGTTACTTTGCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC 699
QY 481 AGCGCATTCACCTGTGCAACAAACCGCTCAGAGTACCTGAAGATATATGTGGCCGC 540
DB 700 AGCGCATTCACCTGTGCAACAAACCGCTCAGAGTACCTGAAGATATATGTGGCCGC 759
QY 541 AGCCCCCTGCTCCCGCGCACACAACATGTGACACAGCAGGTTCCAGGAGTACCTGCGAC 600
DB 760 AGCCCCCTGCTCCCGCGCACACAACATGTGACACAGCAGGTTCCAGGAGTACCTGCGAC 819
QY 601 CGCCAGGGCGGTGGCGCGCTGTGGGGCAGCGTCATCTACCCCTCCATGAAGAAGCCATC 660
DB 820 CGCCAGGGCGGTGGCGCGCTGTGGGGCAGCGTCATCTACCCCTCCATGAAGAAGCCATC 879
QY 661 GCCCAGCCCATCAAGGTGGCCCGCAGGACCACTGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 720
DB 880 GCCCAGCCCATCAAGGTGGCCCGCAGGACCACTGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 939
QY 721 TACGGGGCTGACTTCTCGTCTCTGGAGGAGTTCAGGCGCTTGCTGATCGAGATCAATTC 780
DB 940 TACGGGGCTGACTTCTCGTCTCTGGAGGAGTTCAGGCGCTTGCTGATCGAGATCAATTC 999
QY 781 AGCCCCCACCATGACACCGCTCCACGCGGTTCACGCGCCAGCTGTGTCACAGGTGCGAGGAC 840
DB 1000 AGCCCCCACCATGACACCGCTCCACGCGGTTCACGCGCCAGCTGTGTCACAGGTGCGAGGAC 1059
QY 841 GACACATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCAGCTCTCTGTGG 900
DB 1060 GACACATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCAACTTCAGCTCTCTGTGG 1119
QY 901 AGGCGAG 906
DB 1120 AGCCGG 1125

RESULT 4
ADJ93363
ID ADJ93363 standard; DNA, 3465 BP.
XX
AC ADJ93363;
XX
DT 06-MAY-2004 (first entry)
XX
XX Human BGS-42 protein-related DNA clone C SeqID11.
DE
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; neurotropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary cancer; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
OS Homo sapiens.
XX
XX WO2004/005487-A2.
XX

PD 15-JAN-2004.
XX
PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Feder JN, Wu S, Nelson TC;
XX
DR WPI; 2004-099381/10.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Example 4; SEQ ID NO 11; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytosolic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
XX clone sequence which is related to the invention.
XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Query Match 84.1%; Score 762; DB 12; Length 3465;
Best Local Similarity 92.6%; Pred. No. 9.1e-173;
Matches 839; Conservative 0; Mismatches 0; Indels 67; Gaps 1;

QY 1 GAGGACATCGACAGCTCAGCAGATGCGGTGGAGGACCTCACTGAGGCGGAGTGGAGGAC 60
DB 1894 GAGGACATCGACAGCTCAGCAGATGCGGTGGAGGACCTCACTGAGGCGGAGTGGAGGAC 1953
QY 61 CTGACCCAGCAGTACTACTCCCTGGTTTCATGGCGATCTTCAATTCAGAAAT 120
DB 1954 CTGACCCAGCAGTACTACTCCCTGGTTTCATGGCGATCTTCAATTCAGAAAT 2013
QY 121 TACTTTTCGAGTCCGAGGCTCTGCTGAATAGATCACTGTGTGAACCTCAGACGGAC 180
DB 2014 TACTTTTCGAGTCCGAGGCTCTGCTGAATAGATCACTGTGTGAACCTCAGACGGAC 2073
QY 181 ATTGACGGGTCCGGAACATCTGATATATAAGCCGCGCAAGTCCCGGGGCCGAGAC 240
DB 2074 ATTGACGGGTCCGGAACATCTGATATATAAGCCGCGCAAGTCCCGGGGCCG--- 2129
QY 241 ATAGTGTGTCATGGACCGCTGTGGAGGAGATCTCTGGAGTGGCAGTGCAGACCCCTCTT 300
DB 2130 ----- 2129
QY 301 TCCAGGAGAACAAAGTGGGTGGTCCAGAACTACATCGAGAGCGCGTGTCTCTGTGTGAC 360

Db 2130 ---AGGACAAAGTGGGTGGTCCAGAGTACATCGAGAGCGCGTCTCATCTGTGAC 2186
QY 361 ACCAAGTTCGACATCAGACAGTGGTTCTCTCGTCA CGGACTGGAAACCCCTTGACCATCTGG 420
Db 2187 ACCAAGTTCGACATCAGACAGTGGTTCTCTCGTCA CGGACTGGAAACCCCTTGACCATCTGG 2246
QY 421 TTCTACAGGAGAGTACTTTCGGGTTCTCAACTCAGGCGTCTCCCTGGCAAGCTGGAC 480
Db 2247 TTCTACAGGAGAGTACTTTCGGGTTCTCAACTCAGGCGTCTCCCTGGCAAGCTGGAC 2306
QY 481 AGGCGCATCCACCTGTGCAACACGCGCTCCAGAGTACCTGAAGATGATGTGGGCGCG 540
Db 2307 AGGCGCATCCACCTGTGCAACACGCGCTCCAGAGTACCTGAAGATGATGTGGGCGCG 2366
QY 541 AGCCCTCTGTGCTCCCGCACACAATGTGTGACACAGCACAGGTTCCAGGAGTACCTGCAG 600
Db 2367 AGCCCTCTGTGCTCCCGCACACAATGTGTGACACAGCACAGGTTCCAGGAGTACCTGCAG 2426
QY 601 CGCAGGCGCGTGGCGCGCTGTGGGCGAGGTCATCTACCGTCCATGAGNAGGCCATC 660
Db 2427 CGCAGGCGCGTGGCGCGCTGTGGGCGAGGTCATCTACCGTCCATGAGNAGGCCATC 2486
QY 661 GCCACGCGCATGAAGTGGCGCCAGGACCACTGGAGCTCGCAAGAACAGCTTTGAGCTC 720
Db 2487 GCCACGCGCATGAAGTGGCGCCAGGACCACTGGAGCTCGCAAGAACAGCTTTGAGCTC 2546
QY 721 TAGCGGCGTGAATTCCTTCTTGGAGGAGCTTCAGGCGCGCTGATTCGAGATCAATTC 780
Db 2547 TAGCGGCGTGAATTCCTTCTTGGAGGAGCTTCAGGCGCGCTGATTCGAGATCAATTC 2606
QY 781 AGCCCGCATGACCGCTTCAGCGGTCAGCGGTCAGCGGCGCGCTGTGTCAGAGTGCAGGAG 840
Db 2607 AGCCCGCATGACCGCTTCAGCGGTCAGCGGTCAGCGGCGCGCTGTGTCAGAGTGCAGGAG 2666
QY 841 GACACCATCAAGTGGCGCGCTGGAGCCGAGCTGTGACATCGGCAACTTCGAGCTCTCTGG 900
Db 2667 GACACCATCAAGTGGCGCGCTGGAGCCGAGCTGTGACATCGGCAACTTCGAGCTCTCTGG 2726
QY 901 AGGCAG 906
Db 2727 AGGCAG 2732

RESULT 5
ADJ93361
ID ADJ93361 standard; DNA; 1939 BP.
XX
AC ADJ93361;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 protein-related DNA clone A SeqID9.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiaesthetic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX
OS Homo sapiens.
XX
PN WO2004005487-A2.
XX

PD 15-JAN-2004.
 XX 09-JUL-2003; 2003WO-US021605.
 XX 09-JUL-2002; 2002US-0394725P.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Feder JN, Wu S, Nelson TC;
 XX WPI; 2004-099381/10.
 XX
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 XX useful for preventing, treating or ameliorating a medical condition, e.g.
 XX PT aberrant cellular proliferation, reproductive disorders or testicular
 XX disorders.
 XX
 XX Example 4; SEQ ID NO 9; 343pp; English.
 XX
 XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
 XX be useful for the development of compounds with a cytotatic, respiratory
 XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 XX anti-inflammatory, anabolic, hypertensive, osteopathic, neurotropic,
 XX antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 XX immunosuppressive, antiseborrheic or dermatological activity acting as
 XX tyrosine ligase modulators. In addition, the disclosed sequences may be
 XX useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 XX used for diagnosing a pathological condition or a susceptibility to a
 XX pathological condition in a subject, and for preventing, treating or
 XX ameliorating a medical condition, such as a disorder related to aberrant
 XX tubulin ligase activity, a disorder related to aberrant tubulin-
 XX carboxypeptidase activity, aberrant cellular proliferation, reproductive
 XX disorders, testicular disorders, testicular cancer, pulmonary disorders,
 XX lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 XX neural disorders, brain cancer, liver cancer, or proliferative condition
 XX of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 XX polypeptide, polynucleotide, or their modulators are also useful for
 XX treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 XX disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 XX -42 polypeptide can be used as a preventive agent for immunological
 XX disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 XX disease or scleroderma. The antibodies may be used to purify, detect and
 XX target the BGS-42 polypeptides. The present sequence is that of a DNA
 XX clone sequence which is related to the invention.
 XX
 XX Sequence 1939 BP; 421 A; 592 C; 589 G; 337 T; 0 U; 0 Other;
 XX
 XX Query Match 83.9%; Score 760.4; DB 12; Length 1939;
 XX Best Local Similarity 92.5%; Pred. No. 1.9e-172;
 XX Matches 838; Conservative 0; Mismatches 1; Indels 67; Gaps 1;
 XX
 XX 1 GAGGACATCGACAGTCTACAGAGTCCGCTGGAGGACCTCACTGAGGCCGAGTGGAGGAC 60
 XX |||||
 XX 369 GAGGACATCGACAGTCTACAGAGTCCGCTGGAGGACCTCACTGAGGCCGAGTGGAGGAC 428
 XX |||||
 XX 61 CTGACCCAGCAGTACTCTCTCGTTTCATGGCGATCTTTCATCTCCAAATTCAGAAAT 120
 XX |||||
 XX 429 CTGACCCAGCAGTACTCTCTCGTTTCATGGCGATCTTTCATCTCCAAATTCAGAAAT 488
 XX |||||
 XX 121 TACTTTTCGAGTGCAGGCTCTGTGTAATAGAAATACAGTCTGTGAACCTTCAGACGGAC 180
 XX |||||
 XX 489 TACTTTTCGAGTGCAGGCTCTGTGTAATAGAAATACAGTCTGTGAACCTTCAGACGGAC 548
 XX |||||
 XX 181 ATTGAGGGCTCCGAGATCTGGATTATAAGCCCGCGGCAAGTCCCGGGCCGAGAC 240
 XX |||||
 XX 549 ATTGAGGGCTCCGAGATCTGGATTATAAGCCCGCGGCAAGTCCCGGGCCG---- 604
 XX |||||
 XX 241 ATAGTGTGCATGACCGTGTGGAGGAGATCCTGGAGCTGGCAGTGCAGACCAACCTCTT 300
 XX |||||
 XX 605 ----- 604
 XX 301 TCCAGGGACAAAGTGGGTGTCCAGAAAGTACATCGAGACGCCGCTGCTCATCTGTGAC 360
 XX |||||

Db 605 ---AGGGACAACAAGTGGGTGGTCCAGAGTAGTACATCGAGACGCCGCTGCTCATCTGTGAC 661
 QY 361 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCATCTGTG 420
 Db 662 ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCATCTGTG 721
 QY 421 TTCTACAAGGAGAGTTTACTTTGCGGTTTCTCAACTCAGCGCTTCTCCTGTGACAAAGCTGGAC 480
 Db 722 TTCTACAAGGAGAGTTTACTTTGCGGTTTCTCAACTCAGCGCTTCTCCTGTGACAAAGCTGGAC 781
 QY 481 AGCGCATTCACCTGTGCAACAAACCGCTTCAGAGTACCTGAAAGATGATGTGGGCCCG 540
 Db 782 AGCGCATTCACCTGTGCAACAAACCGCTTCAGAGTACCTGAAAGATGATGTGGGCCCG 841
 QY 541 AGCCCCCTGCTGCCCGCCGACACAAACATGTGGACACAGCAGGTTCAGGAGTACCTGCGAG 600
 Db 842 AGCCCCCTGCTGCCCGCCGACACAAACATGTGGACACAGCAGGTTCAGGAGTACCTGCGAG 901
 QY 601 CGCCAGGGCCGTGGCGCCGCTGTGGGGCAGCGTCATCTACCCGTCCATGAAGAAGGCCATC 660
 Db 902 CGCCAGGGCCGTGGCGCCGCTGTGGGGCAGCGTCATCTACCCGTCCATGAAGAAGGCCATC 961
 QY 661 GCCCAGCCCATGAAGTGGCCCGAGGACACAGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 720
 Db 962 GCCCAGCCCATGAAGTGGCCCGAGGACACAGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 1021
 QY 721 TACGGGGCTGACTTCGCTCTTGGGAGGAGTTCAGGCCCTGGCTGATCGAGATCAATTC 780
 Db 1022 TACGGGGCTGACTTCGCTCTTGGGAGGAGTTCAGGCCCTGGCTGATCGAGATCAATTC 1081
 QY 781 AGCCCCACCATGACACCCCGTCCACGCGGCTCAGCGCCAGCTGTGTGCACAGGTGCGAGGAG 840
 Db 1082 AGCCCCACCATGACACCCCGTCCACGCGGCTCAGCGCCAGCTGTGTGCACAGGTGCGAGGAG 1141
 QY 841 GACACCATCAAGTGGCGCTGACCGAGCTGTGTGACATCGGCAACTTCGAGCTCTCTGTGG 900
 Db 1142 GACACCATCAAGTGGCGCTGACCGAGCTGTGTGACATCGGCAACTTCGAGCTCTCTGTGG 1201
 QY 901 AGGCGAG 906
 Db 1202 AGGCGAG 1207
 XX
 XX RESULT 6
 XX ADJ93359
 XX ID ADJ93359 standard; cDNA; 726 BP.
 XX AC ADJ93359;
 XX DT 06-MAY-2004 (first entry)
 XX DE Human BGS-42 cDNA sequence SeqID3.
 XX KW testis-specific tubulin tyrosine-ligase-like polypeptide;
 XX BGS-42 polypeptide; cytotatic; respiratory-Gen; gastrointestinal-Gen;
 XX neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 XX osteopathic; neurotropic; antiparkinsonian; antiarthritic; antiasthmatic;
 XX anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 XX dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 XX tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 XX testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 XX gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 XX brain cancer; liver cancer; proliferative condition; testis; lung;
 XX small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 XX emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 XX Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 XX sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 XX CDS 1..726

/*tag= a
/product= "Human BGS-42 protein"
/partial
/note= "No start or stop codon"

WO2004005487-A2.

15-JAN-2004.

09-JUL-2003; 2003WO-US021605.

09-JUL-2002; 2002US-0394725P.

(BRIM) BRISTOL-MYERS SQUIBB CO.

Feder JN, Wu S, Nelson TC;

WPI; 2004-099381/10.

New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, useful for preventing, treating or ameliorating a medical condition, e.g. aberrant cellular proliferation, reproductive disorders or testicular disorders.

Disclosure; SEQ ID NO 3; 343pp; English.

This invention relates to a novel testis-specific tubulin tyrosine-ligase-like polypeptide, designated the BGS-42 polypeptide. The invention may be useful for the development of compounds with a cytostatic, respiratory

-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, antiinflammatory, anabolic, hypertensive, osteopathic, nootropic, antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial, immunosuppressive, antiseborrheic or dermatological activity acting as tyrosine ligase modulators. In addition, the disclosed sequences may be useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of a partial cDNA which encodes the human BGS-42 protein of the invention.

Sequence 726 BP; 157 A; 227 C; 214 G; 128 T; 0 U; 0 Other;

Query Match 80.1%; Score 726; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 2.6e-164;
Matches 726; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 181 ATTGACGGGCTCCGGAACATCTGATATATAAGCCGCGGCCAAGTCCCGGGCCGAGAC 240
DB 1 ATTGACGGGCTCCGGAACATCTGATATATAAGCCGCGGCCAAGTCCCGGGCCGAGAC 60
QY 241 ATAGTGTGCATGGAACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCACTCTT 300
DB 61 ATAGTGTGCATGGAACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCACTCTT 120
QY 301 TCAGGAGCAACAAAGTGGGTGTCAGAAATGATATGAGAGCGCGCTGCTATCTGTGAC 360
DB 121 TCAGGAGCAACAAAGTGGGTGTCAGAAATGATATGAGAGCGCGCTGCTATCTGTGAC 180
QY 361 ACCAAGTTCACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCACTGG 420

Db 181 ACCAAGTTCACATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCATCTGG 240
QY 421 TTCTACAGGAGAGTACTTTCGGTTCCTAACTCAGCGCTTCTCCCTGGGAGCAAGCTGAC 480
Db 241 TTCTACAGGAGAGTACTTTCGGTTCCTAACTCAGCGCTTCTCCCTGGGAGCAAGCTGAC 300
QY 481 AGGCGCATCCACCTGTGCAACAAACCGCTCCAGAGTACCTGAAGAATGATGTGGCGGC 540
Db 301 AGGCGCATCCACCTGTGCAACAAACCGCTCCAGAGTACCTGAAGAATGATGTGGCGGC 360
QY 541 AGCCCCCTGCTGCCCGCACACAACATGTGGAGCCAGCAACAGGTTCACAGGAGTACCTGCAG 600
Db 361 AGCCCCCTGCTGCCCGCACACAACATGTGGAGCCAGCAACAGGTTCACAGGAGTACCTGCAG 420
QY 601 CGCAGAGCGCGTGGCGCGCTGTGGGAGCGCTATCTACCCGCTCCATGAAGAAGGCCATC 660
Db 421 CGCAGAGCGCGTGGCGCGCTGTGGGAGCGCTATCTACCCGCTCCATGAAGAAGGCCATC 480
QY 661 GCCCAGCCCATGAAGTGGCCAGGACCACTGCGAGCTCGCAAGAACAGCTTTGAGCTC 720
Db 481 GCCCAGCCCATGAAGTGGCCAGGACCACTGCGAGCTCGCAAGAACAGCTTTGAGCTC 540
QY 721 TAGCGGCTGACTTCTGCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTC 780
Db 541 TAGCGGCTGACTTCTGCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTC 600
QY 781 AGCCCCCATGATGACCCGCTCCAGCCGCTCAGCCGCTCAGCCGCTGTCAGAGTGCAGGAG 840
Db 601 AGCCCCCATGATGACCCGCTCCAGCCGCTCAGCCGCTCAGCCGCTGTCAGAGTGCAGGAG 660
QY 841 GACACCATCAAGTGGCGCTGGAGCCAGCTGTGACATCGGCAACTTCGAGCTCCTCTGG 900
Db 661 GACACCATCAAGTGGCGCTGGAGCCAGCTGTGACATCGGCAACTTCGAGCTCCTCTGG 720
QY 901 AGGCGAG 906
Db 721 AGGCGAG 726

RESULT 7

ACH91699/c

ID ACH91699 standard; DNA; 490 BP.

AC ACH91699;

XX 29-JUL-2004 (first entry)

XX Human genome derived single exon probe #24894.

DE Human; probe; ss; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.

XX Homo sapiens.

XX US2003194704-A1.

XX 16-OCT-2003.

XX 03-APR-2002; 2002US-00029386.

XX 03-APR-2002; 2002US-00029386.

XX (PENN/) PENN S G.

XX (RANK/) RANK D R.

XX (HANZ/) HANZEL D K.

XX Penn SG, Rank DR, Hanzel DK;

XX WPI; 2004-119264/12.

XX New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for

PT surveying tissues.
XX Claim 1; SEQ ID NO 24894; 80pp; English.
XX
CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subscription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterising
CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ Sequence 490 BP; 77 A; 150 C; 168 G; 95 T; 0 U; 0 Other;

Query Match 47.1%; Score 427; DB 12; Length 490;
Best Local Similarity 100.0%; Pred. No. 1.4e-92;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 CAGCGCCATCCACCTGTGCAACACCGCGTCCAGAGTACCTGGAAGATGATGTGGCGC 539
DB |||||||
QY 427 CAGCGCCATCCACCTGTGCAACACCGCGTCCAGAGTACCTGGAAGATGATGTGGCGC 368
DB |||||||

QY 540 CAGCCCCCTGTGCTGCCGCGCACAAACATGTGGACGACGAGTTCAGAGTACCTGCA 599
DB |||||||

DB 367 CAGCCCCCTGTGCTGCCGCGCACAAACATGTGGACGACGAGTTCAGAGTACCTGCA 308
QY 600 GCGCCACGCGCGTGGCGCGTGTGGGCGACGCTCATCTACCGTCCATGAAGAAGCCAT 659
DB |||||||

DB 307 GCGCCACGCGCGTGGCGCGTGTGGGCGACGCTCATCTACCGTCCATGAAGAAGCCAT 248
QY 660 CCGCCACGCGCATGAGGTGGCGCGCGGACGACGAGTTCGCAAGACGCTTTCAGCT 719
DB 247 CCGCCACGCGCATGAGGTGGCGCGCGGACGACGAGTTCGCAAGACGCTTTCAGCT 188
QY 720 CTACGGGGGTGACTTCGTCTCTGGGAGGACTTCAGGCGCTGGCTGATCGAGATCAATTC 779
DB 187 CTACGGGGGTGACTTCGTCTCTGGGAGGACTTCAGGCGCTGGCTGATCGAGATCAATTC 128
QY 780 CAGCCCCACCATGCAACCCGCTCAACGCGGCTCAAGGCGGCGGAGTGTGTGCAAGGTGAGGA 839
DB 127 CAGCCCCACCATGCAACCCGCTCAACGCGGCTCAAGGCGGCGGAGTGTGTGCAAGGTGAGGA 68
QY 840 GGCACACATCAGGTGGCGCGTGGACGCGCTGTGATCGGCAACTTCGAGCTCCTGTG 899
DB 67 GGCACACATCAGGTGGCGCGTGGACGCGCTGTGATCGGCAACTTCGAGCTCCTGTG 8

QY 900 GAGGCAG 906
DB |||||||
DB 7 GAGGCAG 1

RESULT 8
ADQ17814/C
ID ADQ17814 standard; DNA; 101270 BP.
XX
AC ADQ17814;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 631.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
DB
XX Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PX (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 631; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 101270 BP; 24151 A; 28044 C; 28100 G; 20975 T; 0 U; 0 Other;

Query Match 47.1%; Score 427; DB 12; Length 101270;
Best Local Similarity 100.0%; Pred. No. 6.1e-92;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 CAGCGCCATCCACCTGTGCAACACCGCGTCCAGAGTACCTGGAAGATGATGTGGCGC 539
DB |||||||
DB 44821 CAGCGCCATCCACCTGTGCAACACCGCGTCCAGAGTACCTGGAAGATGATGTGGCGC 44762

QY 540 CAGCCCCCTGTGCTGCCGCGCACAAACATGTGGACGACGAGTTCAGAGTACCTGCA 599
DB |||||||

DB 44761 CAGCCCCCTGTGCTGCCGCGCACAAACATGTGGACGACGAGTTCAGAGTACCTGCA 44702

QY 600 GCGCCACGCGCGTGGCGCGTGTGGGCGAGCTCATCTACCGTCCATGAAGAAGCCAT 659
DB |||||||

DB 44701 GCGCCACGCGCGTGGCGCGTGTGGGCGAGCTCATCTACCGTCCATGAAGAAGCCAT 44642

QY 660 CGCCACGCGCATGAAGTGGCCAGGACCAAGTGGAGCTCGCAAGACAGCTTTGAGCT 719
DB 44641 CGCCACGCGCATGAAGTGGCCAGGACCAAGTGGAGCTCGCAAGACAGCTTTGAGCT 44582
QY 720 CTACGGGGCTGACTTCCTCTTGGAGGGAGCTTCAGGCGCCCTGGCTGATCGAGATCAATTC 779
DB 44581 CTACGGGGCTGACTTCCTCTTGGAGGGAGCTTCAGGCGCCCTGGCTGATCGAGATCAATTC 44522
QY 780 CAGCCCCACCATGACCCGCTCCAGCCGGTCAAGCCGGTCAAGCCAGCTGTGTGCAAGGTCAGGA 839
DB 44521 CAGCCCCACCATGACCCGCTCCAGCCGGTCAAGCCGGTCAAGCCAGCTGTGTGCAAGGTCAGGA 44462
QY 840 GGACACCATCAAGTGGCCGCTGGAGCCGACCTGTGACATCGGCAATTCGAGCTCCCTGTG 899
DB 44461 GGACACCATCAAGTGGCCGCTGGAGCCGACCTGTGACATCGGCAATTCGAGCTCCCTGTG 44402
QY 900 GAGGCAG 906
DB 44401 GAGGCAG 44395
RESULT 9
AAC77214
ID AAC77214 standard; cDNA; 2881 BP.
AC AAC77214;
XX 08-FEB-2001 (first entry)
XX Human ORFX ORF2769 polynucleotide sequence SEQ ID NO:5537.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX Homo sapiens.
XX OS
XX PN WO200058473-A2.
XX PD
XX PF 05-OCT-2000.
XX PF 31-MAR-2000; 2000WO-US008621.
XX PR 31-MAR-1999; 99US-0127607P.
XX PR 02-APR-1999; 99US-0127636P.
XX PR 05-APR-1999; 99US-0127728P.
XX PR 30-MAR-2000; 2000US-00540763.
XX (CURA-) CURAGEN CORP.
XX PA
XX PI Shimkets RA, Leach M;
XX DR WPI; 2000-602362/57.
XX DR P-PSDB; AAB43005.
XX DR
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX PS Claim 5; Page 4718-4720; 5507pp; English.
XX

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 2881 BP; 640 A; 812 C; 840 G; 589 T; 0 U; 0 Other;
Query Match 44.2%; Score 400.4; DB 3; Length 2881;
Best Local Similarity 66.1%; Pred. No. 5.5e-86;
Matches 578; Conservative 0; Mismatches 296; Indels 0; Gaps 0;
QY 3 GGACATCGACACGTCAGCAGATCCGCTGGAGGACCTCACTCAGGCCCGAGTGGAGGACCT 62
DB 894 GGACATCGACACGACCTGGAGGCCCGCTGTACTCACCCTCCGAGGGCTGGTCCCTCTT 953
QY 63 GACCCAGAGTACTACTCCCTCTGATGCGGATGCTTTTCATCTCCAAATCAAGAAATTA 122
DB 954 CCTCCAGCGCTACTACCAAGTGTGCCAAGGGGAGAACTCAGGCACCTCGACACTCA 1013
QY 123 CTTTTCGAGTGCAGGCTCTGCTGTAATGAATCAGCTGTGAACCTCAGACGGACAT 182
DB 1014 GTTCAGCGCTGTGAGGACATCTCGACGAGCTGAGCGCGTGTATCCCGATAGACAT 1073
QY 183 TGACGGGCTCCGGAACATCTGGATTATAAGCCCGGCCCAAGTCCCGGGCCGAGACAT 242
DB 1074 GGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGACCAAGTCCCGTGGACGAGCAT 1133
QY 243 AGTGTGATGAGACCGTGTGGAGGAGATCTCGAGCTGGAGCTGGACACCCCTCTTTC 302
DB 1134 CATGTGATGAGACCACTGGAGGAGATGCTGAAGCTGTGAAGCCGCAACCCGCTGTGAT 1193
QY 303 CAGGGACAACAGTGGTGTGTCAGAGTATCAGAGCCGCTGCTCATCTGTGACAC 362
DB 1194 GAAGGACGGCAAGTGGTGTGAGAGTATATGAGCGGCCCTCTCTCATCTTTGGCAC 1253
QY 363 CAAAGTTGACATCAGACAGTGGTTCCTCTGTCAGGAGCTGGAACCCCTGACCATCTGGTT 422
DB 1254 CAAAGTTGACCTCAGACAGTGGTTCCTGTTAATGACTGGAACCCACTTACCGTGTGTT 1313
QY 423 CTACAAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGAACAAGCTGGACAG 482
DB 1314 CTACCCGCGCAGCTATATCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCCTGGCAA 1373
QY 483 CGCCATCCACCTGTGCAACAACCGCTCCAGAGTACCTGGAAGATGATGTGGGCCGACAG 542
DB 1374 CTCAGTGCACCTGTGCAACAACCTCCATCCAGAGACCTTGGAGAACTCATGCCATCGGCA 1433
QY 543 CCCCCTGCTGCCCGCACACAACATGTGGACCAACAGAGTTCAGGAGTACCTGACGCG 602
DB 1434 TCCACTGCTTCCGCCAGACACAATGTGTGTACCCAGAGGTTCCAGGCCCACTTGCAGGA 1493
QY 603 CCAGGCGCTGTGGCGCGCTGTGGGCGAGCGTCACTATCCCGTCCATGAAGAAGCCATCGC 662
DB 1494 GATGGGTGCCCCAAATGCTGTGGTCCACCATCATCGTCCCTGGCATGAAGGATGCTGTGAT 1553
QY 663 CCAGGCCATGAAGTGGCCCCCAGGACCAACAGCTGGAGCCCTCGCAAGACAGCTTTGAGCTTA 722

Db	1554	CCAGCACTTCAGACTCCACAGGACACCGTCAGTGTCCGAAGCCAGCTTTGAGCTCTA	1613
Qy	723	CGGGCTGACTTTCGTCTCTTGGAGGAGACTTCAGGCCCCCTGGCTGATCCAGATCAATTCAG	782
Db	1614	TGGCGCTGACTTTCGTGTTCGGGGAGGAGCTTCAGGCCCTGGCTGATTGAGATCAAGCCAG	1673
Qy	783	CCCCACCATGCACCCGCTCCACGCCGGTCA CGGCCCCAGCTGTGTGTCACAGGTGCAGGAGGA	842
Db	1674	CCCCACGATGGCACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGC AAGTGA	1733
Qy	843	CACCATCAAGGTGGCCGTGGACCCGACGCTGTGAC	876
Db	1734	CACCTCGCGGTGCTGATTCATTGACCGGAGGCTGGAC	1767

RESULT 10

RESOLUT
AAH16735

AAH16735
ID AAH16735 standard; cDNA; 2326 BP.

XX
XX
CC/OT/000000

AAH16735;

XX
XX
/CC/OTR/

DT 26-JUN-2001 (first entry)

XX

Human CDNA sequence SEQ ID NO:15920.

XX XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapi

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-20

XX

PR 29-JUL-19

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-20
PR 02-MAY-20

PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JB-00241888

PR 09-JUN-20
XX

XX
PA (HELT-) HELIX RES INSTXX
FA (HELI-7) R

PI Ota T. Isogai T. Nishikawa

PI Ishii S,

XX
10111111 0, 0ag+yma 1,

DR WPI; 2001

XX XX

PT Primer se

PT length CD

PT diagnosis

PT cDNAs.

XX

PS Claim 8;

XX

The present

(a) an oligo-dT primer and an oligonucleotide complementary to the length cDNAs defined in the specification, where a primer set comprises a complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly, full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

RESULT 11

CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB924446 to AAB958933	
CC	represent human amino acid sequences; and AAH13629 to AAH13632 represent	
CC	oligonucleotides, all of which are used in the exemplification of the	
CC	present invention	
XX		
SQ	Sequence 2326 BP; 524 A; 701 C; 642 G; 459 T; 0 U; 0 Other;	
	Query Match 44.0%; Score 398.8; DB 4; Length 2326;	
	Best Local Similarity 66.3%; Pred. No. 1.3e-85;	
	Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;	
QY	3 GGACATCGACACGTACACAGATGCCGTGGAGGACCTCACTAGGCCCGAGTGGGAGNACCT 62	
DB		
DB	732 GGACATCGACAAAGAACCTGGAGGCCCGCGTGTACCTCACCCCGAGGGCGTGGTCCCTCT 791	
QY	63 GAACCCACGCTACTCTCCCTCGTTATGCGCATGCTTTTCATCTCCAATTCAGAAATTA 122	
DB		
DB	792 CCTCCAGCGCTACTACCAAGTGTTCACGAAAGGGGAGAACTCAGGCACCTCGACACTCA 851	
QY	123 CTTTTCGACGTCCGAGCTCTCGTGAATAGAATCAACGTCTGTGAACCTTCAGACGGACAT 182	
DB		
DB	852 GGTCCAGCGCTGTGAGGACATCTCGACGAGCTGCAGGCCGTGTGATCCCAATAGACAT 911	
QY	183 TGACGGGCTCCGAAACATCTGGATTATAAGCCCGGGCCAAAGTCCCGGGCGCGAGACAT 242	
DB		
DB	912 GGAAGGGGATCGCAACATCTGSAATCTGTGAAGCAGGAGCCAAAGTCCCGCGGACGAGCAT 971	
QY	243 AGTGTGATCGACCGTGTGACGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTTC 302	
DB		
DB	972 CATGTGATCGACCACTGGAGAGATGCTGGAAGCTGTGAAACGGCAACCCCGTGTGAT 1031	
QY	303 CAGGGAACAAGTGGTGGTCCAGAAAGTACATCGAGACGCCGCTCTCATCTGTGACAC 362	
DB		
DB	1032 GAAGGACGGCAAGTGGTGGTGCAGAAAGTATAATTGAGCGGCCCTCTCTCATCTTTGGCAC 1091	
QY	363 CAAGTTCGACATCAGACAGTGGTTCTCTGTCACGAGATGGAACCCCTGACCATCTGGTT 422	
DB		
DB	1092 CAAGTTTGACCTCAGACAGTGGTTCTCTGSGTAACTGACTGGAACCCCACTTACCGTGTGGTT 1151	
QY	423 CTACAGAGGAGTTACTTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAG 482	
DB		
DB	1152 CTACCGCGACAGCTATATTCGCTTTTCCACGACGCCCTTCTCCCTGAAGAACTGATCGCCATCGGCA 1211	
QY	483 CGCCATCCACTGTGCACAAACCCCGTCCAGAAAGTACCTGAAGAATGATGTGGGCCGACAG 542	
DB		
DB	1212 CTCACTGCACTGTGCACAACTCCATCCAGAAAGCACTGGAGAACTCATGTCCATCGGCA 1271	
QY	543 CCCCTGCTCCCGCGCACAAACATGTGGACCAAGCCAGTCCAGAGGTACCTGCGAGCG 602	
DB		
DB	1272 TCCACTGCTTCCGCCAGACAACTATGTGTTCTAGCCAGAGGTTCCAGGCCCACTGCGAGGA 1331	
QY	603 CCAGGGCCGTGGCGCGTGTGGGCGACGCTCATCTACCCCTCCATGAAGAAGGCCATCGC 662	
DB		
DB	1332 GATGGGTGCCCCAATGCTTGGTTCACCATCATCTGTCCTGGCATGAAGATGCTGTGAT 1391	
QY	663 CCAACGATGAAGGTGGCCCAAGACCAAGTGGAGCCTTCGCAAGAAACAGCTTTGAGCTCTA 722	
DB		
DB	1392 CACGCACTTCAGACCTCCCAGGACACCGGTGCAGTGTGGAAAGCCAGCTTTGAGCTCTA 1451	
QY	723 CGGGGTGACTTTCGTCTTGGGAGGACCTTCAGGCCCTGGCTCATCGAGATCAATTCAC 782	
DB		
DB	1452 TGGCGCTGACTTCTGTGTTGGGGAGGACTTCCAGCCCTTGGCTCATGTGAGATCAACGCCAG 1511	
QY	783 CCCACCATGCACCCCGTCCACCGCGGTACGGGCCACAGCTGTGTGCAAGTGTGAGGAGGA 842	
DB		
DB	1512 CCCCAAGTGGACCCCTCCACAGAGTCACTGCCCGGCTCTGTGCTGGCGTGAAGCTGA 1571	
QY	843 CACCATCAAGGTGGCGGTGGACCGCA 868	
DB	1572 CACCTGCGGCTGGTCAATTGACCGGA 1597	

RESULT 11

KW T-cell; immune response.
 XX Homo sapiens.
 XX US2003109434-A1.
 XX 12-JUN-2003.
 XX 19-MAR-2002; 2002US-00102524.
 XX 19-MAR-2001; 2001US-0277245P.
 PR 21-DEC-2001; 2001US-0343340P.
 XX (CORI-) CORIXA CORP.
 XX Algate PA, Mannion J, Gaiger A, Gordon B, Harlocker SL;
 XX WPI; 2002-759855/82.
 XX New isolated polynucleotides and polypeptides, useful for detecting the
 PT presence of, and treating cancer, particularly kidney cancer by
 PT stimulating T-cells specific for a tumor protein, and stimulating immune
 PT response in a patient.
 XX Claim 1; SEQ ID NO 1765; 78pp; English.
 XX The invention relates to a new isolated polynucleotide (a Human kidney
 CC tumour specific cDNA) comprising any one of the 1855 sequences identified
 CC in the specification (or their complements, degenerate variants,
 CC sequences consisting of at least 20 contiguous residues them, sequences
 CC that hybridise to them under highly stringent conditions or sequences
 CC having at least 75 or 90% sequence identity to the 1855 sequences. Also
 CC included are detecting/determining the presence of cancer in a patient,
 CC stimulating an immune response in a patient; treating kidney cancer in a
 CC patient; an isolated polypeptide encoded by one of the 1855 sequences, an
 CC expression vector comprising the polynucleotide operably linked to an
 CC expression control sequence, a host cell transformed/transfected with the
 CC vector, an isolated antibody (or its antigen-binding fragment) that
 CC specifically binds to the protein, a fusion protein comprising at least
 CC one of the proteins, stimulating and/or expanding T-cells specific for a
 CC tumour protein, an isolated T-cell population comprising the T-cells, a
 CC composition comprising a first component (such as a carrier or
 CC immunostimulant) and a second component (comprising one of the
 CC polynucleotides, the polypeptides, an antibody, T-cell or an antigen-
 CC presenting cell that expresses the polynucleotide) and a diagnostic kit
 CC comprising at least one of the oligonucleotides, or at least one antibody
 CC and a detection reagent comprising a reporter group. The polynucleotides,
 CC polypeptides, antibodies and antigen-presenting cells are useful for
 CC detecting the presence of, and treating cancer, particularly kidney
 CC cancer by stimulating and/or expanding T-cells specific for a tumour
 CC protein, and stimulating immune response in a patient. The present
 CC sequence is one of the Human kidney tumour specific cDNAs. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?docID=20030109434.
 XX
 SQ Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;

Query Match 44.0%; Score 398.8; DB 7; Length 2538;
 Best Local Similarity 66.3%; Pred.No. 1.3e-85;
 Matches 5/4; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 3 GGACATCGACAGCTCAGCAGATGCGCTGGAGGACCTCACTGAGGCGGAGTGGAGGACCT 62
 Db |||||
 894 GGACATCGACAGGACCTGGAGGCGCGCTGTACCTCACCCGAGGGCTGTCCCTCT 953
 QY 63 GACCCAGCAGTACTACTCCCTCGTTCACTGGCGATGCTTCACTCCAAATCAAGAAATTA 122
 Db |||||
 954 CCTCCAGCGCTACTACCAAGTGTGTCCACGAAGGGGAGAACTCAGGCACTCGACACTCA 1013
 QY 123 CTTTTCGAGTCCAGGCTCTGTGAATAGATACAGTCTGTGAACCTTCAGACGACAT 182
 Db |||||
 1014 GGTCCAGCCCTGTGAGGACATCTCGAGCAGCTGCAGGCGGTGGTACCCCAATAGACAT 1073

QY 183 TGACGGGCTCGGAAACATCTGGATTATAAGCCCGCGCCAAAGTCCCGGGCGGACAT 242
 Db |||||
 1074 GGAAGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAGTCCCGGGAGGAGCAT 1133
 QY 243 AGTGTGCATGGACCGCTGTGGAGGAGATCTCTGAGCTGGCAGCTGCAGACACCCTCTTTC 302
 Db |||||
 1134 CATGTGCATGGACCACTCTGGAGGAGATGCTGAAGCTGTGTGAACGCGCAACCCCTGTGTGAT 1193
 QY 303 CAGGGACAAAGTGGTGTCCAGAAATGACATCCAGAGCGCGCTGCTCATCTGTGACAC 362
 Db |||||
 1194 GAAGGACGCAAGTGGTGTGCAGAAATATATGAGCGGCCCTCTCTCATCTTTTGGCAC 1253
 QY 363 CAAAGTTCCGACATCAGACAGTGTCTCTCGTCA CGGACTGGAACCCCTGACCATCTGTGTT 422
 Db |||||
 1254 CAAAGTTGACCTCAGACAGTGTCTCTGTGTAACCTGACTGGAACCCCACTTACCGTGTGTT 1313
 QY 423 CTACAGGAGAGTACTTCTGCGGTTCTCAACTCAGCGCTTCTCCCTGGAGCAAGCTGACAG 482
 Db |||||
 1314 CTACCGCGACAGCTATATCCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCTGGACAA 1373
 QY 483 CGCCATCCACCTGTGTCAACAAACGCGTCCAGAAATGACCTGGAAGATGATGTGGCGCGCAG 542
 Db |||||
 1374 CTCAGTGCACCTGTGCAACACTCTCCATCCAGAAAGCACTTGGAGAACTCATGCCATCGGCA 1433
 QY 543 CCCCCTGTGCTCCCGCACCAACATGTGGACCAACAGCTGAGGTTCCAGGAGTACTCTGCAGCG 602
 Db |||||
 1434 TCCACTGTCTCCGCGCAGACAAACATGTGTGTAGCCAGAGGTTCCAGGCCCACTCTGCAGGA 1493
 QY 603 CAGGGCGTGGCGCGGTGTGGGCGAGCTCATCTACCGTCCATGAGAAAGGCCATCGC 662
 Db |||||
 1494 GATGGTGTGCCCCAAATGTCTTGGTCCACCATCATCTGCTGCGCTGCGCATGAAGGATGCTGTGAT 1553
 QY 663 CCACGCCATGAAGTGGCGCCAGGACCACTGAGAGCTCGCAAGAACAGCTTTGAGCTCTA 722
 Db |||||
 1554 CACGCACTTACAGCTTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTA 1613
 QY 723 CGGGGCTGACTTTCGCTCTTGGGAGGAGCTTCAGGCGCTGCTGATCGAGATCAATTCACAG 782
 Db |||||
 1614 TGGCGCTGACTTTCGTTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAAGCCCA 1673
 QY 783 CCCCACCATGACCCGTCACCGCTCAGGCCCAAGCTGTGTGACAGGTGTCAGAGGAGGA 842
 Db |||||
 1674 CCCCAGATGGCACCCTCCACAGCAGTCACTCCCGGCTCTGTGCTGGCGTGCAGAGCTGA 1733
 QY 843 CACCATCAAGTGGCGCGTGGACCGCA 868
 Db |||||
 1734 CACCTGCGGTGTGTCATTGACCGGA 1759

RESULT 13
 ACN37881
 ID ACN37881 standard; cDNA; 2553 BP.
 XX ACN37881;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Tumour-associated antigenic target (TAT) cDNA DNA324273, SEQ ID NO:1057.
 XX
 KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
 KW tumour; diagnosis; cell proliferative disorder; breast cancer;
 KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
 KW central nervous system cancer; bladder cancer; pancreatic cancer;
 KW cervical cancer; melanoma; leukaemia; hybridisation probe;
 KW chromosome identification; chromosome mapping; gene mapping;
 KW gene therapy; cytostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2004030615-A2.
 XX
 PD 15-APR-2004.

XX 29-SEP-2003; 2003WO-US028547.
PF
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
PI Wu TD, Zhang Z, Zhou Y;
XX
DR WPI; 2004-347921/32.
DR P-PSDB; ABM80420.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 1; SEQ ID NO 1057; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
SQ Sequence 2553 BP; 575 A; 716 C; 733 G; 529 T; 0 U; 0 Other;

Query Match 44.0%; Score 398.8; DB 13; Length 2553;
Best Local Similarity 66.3%; Pred. No. 1.3e-85;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 3 GGACATCGACAGCTGACAGATGCGTGGAGGACTCACTGAGCCGAGTGGAGGACCT 62
DB |||||
DB 894 GGACATCGACAGGACCTGGAGGCCCGCTGTACTCACCCCGAGGGCTGTCTCTT 953
QY 63 GACCCAGCAGTACTACTCCCTCGTTCTATGCGGATGCTTTCATCTCCAATTCAGAAATTA 122
DB |||||
DB 954 CCTCCAGCGCTACTACCAAGTGTGTCACGAAGGGGCGAGAACTCAGGCACCTCGACACTCA 1013
QY 123 CTTTTCGAGTGCAGGCTGTCTGTGATAGATACGCTGTGTGACCTCAGACGGACAT 182
DB |||||
DB 1014 GGTCACGCGCTGTGAGGACATCTCTGACGAGCTGTCAGGCGCGTGGTATCCCGATAGACAT 1073
QY 183 TGACGGGCTCGGAACACTCTGGATTATTAAGCCCGCGCAAGTCCCGGGCGCGAGACAT 242
DB |||||
DB 1074 GGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGACCAAGTCCCGCGGACGAGGCAT 1133
QY 243 AGTGTGATGGAACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTTTC 302
DB |||||
DB 1134 CATGTGATGGACCACTGGAGGAGATGCTGAAGCTGTGAACGGCAACCCCGTGGTAT 1193
QY 303 CAGGGACACAGTGGGTGTCGAGAAGTACATCGACAGCCGCTGTCTATCTGTGACAC 362
DB |||||
DB 1194 GAAGACGGCAAGTGGGTGTCAGAAGTATATTGAGCGGCCCTCTCATCTTTGGCAC 1253
QY 363 CAAGTTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGGAACCCCTGACCATCTGGTT 422
DB |||||
DB 1254 CAAGTTTGACCTCAGACAGTGGTTCCTCGTAACTGACTGGAACCCCACTTACCGGTGGTT 1313

QY 423 CTACAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGACAAAGCTGGACAG 482
DB |||||
DB 1314 CTACCGGACAGCTATATCGGCTTTTCCACGACGACCCCTTCTCCCTGAAGAACCTGSAAC 1373
QY 483 CGCATCTCACTGTGCAACAACGCGTCCAGAGTACTCTGAAGAAATGATGTGGCCCGCAG 542
DB |||||
DB 1374 CTGAGTGCACTGTGCAACAACCTCCATCCAGAACACCTGGAGAACTCATGCGCATCGGA 1433
QY 543 CCCCTGCTGCCCGCACACAACATGTGGACCAAGTCCAGGTTCCAGAGTACCTGCGACGC 602
DB |||||
DB 1434 TCCACTGCTTCCGCCAGACAACATGTGCTAGCCAGAGGTTCCAGGCCCACTGCGACGA 1493
QY 603 CAGGCGCGTGGCGCGTGTGGGCGAGCTCATCTACCCCTCCATGAAGAGGCCATCGC 662
DB |||||
DB 1494 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGGATGCTGTGAT 1553
QY 663 CCACGCCATGAAGTGGCGCCAGGACCAACGCTGGAGCCCTCGCAAGAAACAGCTTTGAGCTCTA 722
DB |||||
DB 1554 CCAGGCACCTTCAGACCTCCAGGACACCGTGCAGTGTCCGAAGCCAGCTTTGAGCTCTA 1613
QY 723 CGGGGCTGACTTCTGCTCTTGGGAGGAGCTTTCAGGCCCTGGCTGATCGAGATCAATTCAG 782
DB |||||
DB 1614 TGGCGTCTGACTTCTGCTTTCGGGAGGAGCTTTCAGGCCCTGGCTGATTCAGATCAACGCCAG 1673
QY 783 CCCACCATGCAACCTGTCACGCGGTTCAGCGGCCAGCTGTGTGCACAGGTGCGAGGAGGA 842
DB |||||
DB 1674 CCCACGATGGCAACCTTCCACAGCAGTCACTGCGCCGCTCTGTCTGCTGGCGTGAAGCTGA 1733
QY 843 CACCATCAAGTGGCGCTGGACCGCA 868
DB |||||
DB 1734 CACCTGCGGTGCTATTGACCGGA 1759

RESULT 14
ADM03081
ID ADM03081 standard; cDNA; 2848 BP.
XX
AC ADM03081;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cDNA of the invention SEQ ID NO:1766.
XX
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
DR P-PSDB; ADM05524.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 1766; 305pp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 11:30:53 ; Search time 4205.65 Seconds
(without alignments)
10438.446 Million cell updates/sec

Title: US-10-615-659-1_COPY_369_1274
Perfect score: 906
Sequence: 1 gaggacatgcacagtcagc.....tcgagctctgtggaggcag 906

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	817	90.2	817	6	CQ724907 Sequence
2	427	47.1	101270	9	HS355C18
3	400.4	44.2	4238	9	AL022327 Human DNA
4	398.8	44.0	2326	6	AL833939 Homo sapi
5	398.8	44.0	2326	6	BD158727 Primer fo
6	398.8	44.0	2326	6	AX881015 Sequence
7	398.8	44.0	2326	9	AK023960 Homo sapi
8	398.8	44.0	2380	6	AX301197 Sequence
9	398.8	44.0	2553	9	HS0800637
10	398.8	44.0	2848	6	AX834642 Sequence
11	398.8	44.0	3001	6	AK097236 Homo sapi
12	376.8	41.6	1684	9	AF078842 Sequence
13	366.4	40.4	1897	10	BC006830 Homo sapi
14	281.6	31.1	163783	10	AC119959 Mus muscu
15	281.6	31.1	201420	2	AC117700 Mus muscu
16	274.2	30.3	205949	2	AC134940 Rattus no
17	274.2	30.3	218249	2	AC097425 Rattus no
18	212.2	23.4	5282	6	CQ842940 Sequence
19	212.2	23.4	5282	9	AK125875 Homo sapi

C	20	208.6	23.0	3828	6	CQ850619 Sequence
C	21	208.6	23.0	3828	9	AK127786 Homo sapi
C	22	207.4	22.9	2979	6	CQ595968 Sequence
C	23	192.2	21.2	145435	2	AC026685 Homo sapi
C	24	192.2	21.2	161903	2	AC021996 Homo sapi
C	25	191.4	21.1	1560	9	BC009479 Homo sapi
C	26	190.6	21.0	1958	6	BD160681 Primer fo
C	27	190.6	21.0	1958	6	AX884044 Sequence
C	28	190.6	21.0	1958	9	AK024110 Homo sapi
C	29	190.6	21.0	155313	2	AC068315 Homo sapi
C	30	190.6	21.0	185067	9	AC022382 Homo sapi
C	31	190.6	21.0	189430	2	AC011610 Homo sapi
C	32	190.6	21.0	191834	2	AC026196 Homo sapi
C	33	190.6	21.0	197360	2	AC018829 Homo sapi
C	34	190.6	21.0	216911	9	AC018809 Homo sapi
C	35	181.2	20.0	224077	2	AC120010 Mus muscu
C	36	181.2	20.0	266888	2	AC137881 Homo sapi
C	37	174.2	19.2	2543	6	CQ595656 Sequence
C	38	146.4	16.2	220290	2	BX936393 Danio rer
C	39	145.8	16.1	161903	2	AC021996 Homo sapi
C	40	145.8	16.1	222994	2	AC016947 Homo sapi
C	41	144.8	16.0	4615	6	CQ595655 Sequence
C	42	144.8	16.0	76854	3	AC003052 Drosophil
C	43	144.8	16.0	86398	2	AC017785 Drosophil
C	44	144.8	16.0	167201	3	AC092397 Drosophil
C	45	144.8	16.0	259718	3	AE003614 Drosophil

ALIGNMENTS

RESULT 1
CQ724907
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
ORIGIN

CQ724907
Sequence 10841 from Patent WO02068579.
CQ724907
CQ724907.1 GI:42285764

Homo sapiens (human)
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof

Patent: WO 02068579-A 10841 06-SEP-2002;
PE Corporation (NY) (US)

Location/Qualifiers
1. .817
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

Query Match
Best Local Similarity 90.2%; Score 817; DB 6; Length 817;
Matches 817; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

90 TGGCGATGCTTTTCATCTCCAAATTCAGAAATTTACTTTCCAGTGCAGGCTCTGCTGAA 149
1 TGGCGATGCTTTTCATCTCCAAATTCAGAAATTTACTTTTCAGTGCAGGCTCTGCTGAA 60

150 TAGAATCACGCTGTGAACCTCAGACGACATTGACGGGCTCCGGAACATCTGATTAT 209
61 TAGAATCACGCTGTGAACCTCAGACGACATTGACGGGCTCCGGAACATCTGATTAT 120

210 AAAGCCCGCGGCCAAGTCCCGGGCCGAGACATAGTGTGCATGACCGTGTGGAGGAGAT 269
121 AAAGCCCGCGGCCAAGTCCCGGGCCGAGACATAGTGTGCATGACCGTGTGGAGGAGAT 180

270 CCTGGAGCTGGGAGCTGCAGACCACTCTTTTCAGGGGACCAACAGTGGGTGTGCCAGAA 329

Db	181	CTGGAGCTGGCAGCTGCAGACACACCTCTTTCCAGGGAACAACAGTGGTGCAGAA	240
Qy	330	GTATCATCGAGACGCCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCT	389
Db	241	GTATCATCGAGACGCCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCT	300
Qy	390	CGTACCGAGCTGGAACCCCTGACCATCTGGTTCTACAAGGAGGTACTTGCGGTTCTC	449
Db	301	CGTACCGAGCTGGAACCCCTGACCATCTGGTTCTACAAGGAGGTACTTGCGGTTCTC	360
Qy	450	AACTACGCGTTCTCCCTGACAGCTGACACGCCCATCCACCTGTGCAACACGCCGT	509
Db	361	AACTACGCGTTCTCCCTGACAGCTGACACGCCCATCCACCTGTGCAACACGCCGT	420
Qy	510	CCAGAACTACCTGGAAGATGATGTGGCGCGCAGCCCTGCTGCCCGCACACAAATGTG	569
Db	421	CCAGAACTACCTGGAAGATGATGTGGCGCGCAGCCCTGCTGCCCGCACACAAATGTG	480
Qy	570	GACCAACACAGGTTCCAGGAGTACCTGCAGCGCCAGGGCCGTCGCGCGTGTGGGGCAG	629
Db	481	GACCAACACAGGTTCCAGGAGTACCTGCAGCGCCAGGGCCGTCGCGCGTGTGGGGCAG	540
Qy	630	CGTATCTACCGTCCATGAAGAGCCATCGCCCHACGCCATGAAGTGGGCCAGGACCA	689
Db	541	CGTATCTACCGTCCATGAAGAGCCATCGCCCHACGCCATGAAGTGGGCCAGGACCA	600
Qy	690	CGTGGAGCTCGCAAGACAGCTTTCAGCTCTACGGGGCTGACTTCGCTTGGGAGGCA	749
Db	601	CGTGGAGCTCGCAAGACAGCTTTCAGCTCTACGGGGCTGACTTCGCTTGGGAGGCA	660
Qy	750	CTTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCACCATGCACCCGTCACGCGGT	809
Db	661	CTTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCACCATGCACCCGTCACGCGGT	720
Qy	810	CAGGCGCGAGCTGTGTGCACAGTGTGACAGGACACCATCAAGTGGCGCGTGCACCGCAG	869
Db	721	CAGGCGCGAGCTGTGTGCACAGTGTGACAGGACACCATCAAGTGGCGCGTGCACCGCAG	780
Qy	870	CTGTGACATCGGCAACTTCGAGTCTCTGTGGAGGCGAG	906
Db	781	CTGTGACATCGGCAACTTCGAGTCTCTGTGGAGGCGAG	817

RESULT 2	HS355C18	101270 bp	DNA	linear	PRI 05-JUN-2003
LOCUS	Human DNA sequence from clone RP3-355C18 on chromosome 22q13.3				
DEFINITION	Contains the KIAA0027 gene, ESTs, STSs, GSSs and seven putative CpG islands, complete sequence.				
ACCESSION	AL022327				
VERSION	AL022327.17				
KEYWORDS	HTG; CpG island; KIAA0027.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 101270)				
AUTHORS	Cobley V.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk				
COMMENT	On Jun 30, 1999 this sequence version replaced gi:5262834. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi., EMBL; Swi., SWISSPROT; Tri., TREMBL; Wp., WORMPEP; Information				

on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep ----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk -----	
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22 RP3-355C18 is from the library RPCI-3 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pCYPAC2	
This sequence is the entire insert of clone RP3-355C18 The true left end of clone RP5-898I4 is at 5390 in this sequence.	
FEATURES	Location/Qualifiers
Source	1..101270 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="22" /map="q13.3" /clone="RP3-355C18" /clone_lib="RPCI-3"
misc_feature	1..1038 /note="match: STS: Em:AL022483"
repeat_region	complement(1..242) /note="AluJo repeat: matches 1..237 of consensus"
repeat_region	358..484 /note="L1M4 repeat: matches 829..959 of consensus"
repeat_region	485..792 /note="AluSx repeat: matches 2..305 of consensus"
repeat_region	795..1095 /note="AluSx repeat: matches 1..299 of consensus"
repeat_region	1096..1165 /note="L1M4 repeat: matches 960..1031 of consensus"
repeat_region	1166..1563 /note="WSTB1 repeat: matches 4..432 of consensus"
repeat_region	1564..1750 /note="L1M4 repeat: matches 1031..1203 of consensus"
repeat_region	1757..2038 /note="AluSx repeat: matches 1..282 of consensus"
repeat_region	complement(2054..3342) /note="SVA repeat: matches 1..1386 of consensus"
repeat_region	3343..3359 /note="2.8 copies 6 mer TCTCTT 34% conserved"
repeat_region	3360..3417 /note="LTR39-int repeat: matches 3904..3961 of consensus"
repeat_region	3418..3577 /note="L1M4 repeat: matches 1203..1363 of consensus"
repeat_region	3578..3861 /note="AluJo repeat: matches 1..300 of consensus"
repeat_region	3862..4334 /note="L1M4 repeat: matches 1363..1830 of consensus"
repeat_region	4341..4489 /note="L1M4 repeat: matches 2065..2311 of consensus"
repeat_region	complement(4492..4840) /note="MER41 repeat: matches 96..472 of consensus"
repeat_region	4990..4999 /note="2.5 copies 4 mer CCTG 20% conserved"
repeat_region	complement(5142..5300) /note="AluDb repeat: matches 137..295 of consensus"


```
repeat_region complement(5315..5623)
repeat_region /note="AluX repeat: matches 3..312 of consensus"
repeat_region 5773..5863
repeat_region /notes="MR81 repeat: matches 17..112 of consensus"
repeat_region 5917..5935
repeat_region /notes="2.7 copies 7 mer GGGCCCA 29% conserved"
repeat_region 6130..6140
repeat_region /notes="3.7 copies 3 mer GGT 22% conserved"
repeat_region 6139..6148
repeat_region /notes="2.5 copies 4 mer GGAG 20% conserved"
repeat_region 6229..6240
repeat_region /notes="2.0 copies 6 mer CAGCAC 24% conserved"
repeat_region 6573..6590
repeat_region /notes="3.0 copies 6 mer CAGGAG 27% conserved"
repeat_region 6691..6700
repeat_region /notes="2.0 copies 5 mer TGTGC 20% conserved"
repeat_region 6731..6862
repeat_region /notes="3.3 copies 40 mer
repeat_region CCAGTTAGGAGGACCTGGCGAGTGGCAGCTGCA 219% conserved"
repeat_region 7338..7356
repeat_region /notes="2.7 copies 7 mer GCCCAC 29% conserved"
repeat_region 7636..7664
repeat_region /notes="2.1 copies 14 mer CCCAGAGTGGCC 49% conserved"
repeat_region 7893..7906
repeat_region /notes="2.0 copies 7 mer TGCACC 28% conserved"
repeat_region 7969..7980
repeat_region /notes="2.0 copies 6 mer CCATGA 24% conserved"
repeat_region 8306..8316
repeat_region /notes="2.8 copies 4 mer GTGG 22% conserved"
repeat_region 8399..8414
repeat_region /notes="3.2 copies 5 mer GAGGG 23% conserved"
repeat_region 8840..8851
repeat_region /notes="2.2 copies 5 mer AGCTG 24% conserved"
repeat_region 8939..8958
repeat_region /notes="2.2 copies 9 mer CCCGCCCA 40% conserved"
repeat_region 9125..9135
repeat_region /notes="2.2 copies 5 mer GTCT 22% conserved"
repeat_region 9246..9256
repeat_region /notes="3.7 copies 3 mer GGA 22% conserved"
repeat_region 9263..9274
repeat_region /notes="2.0 copies 6 mer CAGTGA 24% conserved"
repeat_region 9373..9383
repeat_region /notes="2.2 copies 5 mer CCTGA 22% conserved"
repeat_region 9434..9444
repeat_region /notes="2.2 copies 5 mer CAGGC 22% conserved"
repeat_region 9616..9631
repeat_region /notes="2.3 copies 7 mer GGGGATG 32% conserved"
repeat_region 10067..10076
repeat_region /notes="2.5 copies 4 mer GGCA 20% conserved"
repeat_region 10245..10254
repeat_region /notes="2.5 copies 4 mer CCTT 20% conserved"
repeat_region 10496..10531
repeat_region /notes="2.0 copies 18 mer GCGAGGACAGGACAGG 63% conserved"
repeat_region 10498..10531
repeat_region /notes="2.8 copies 12 mer CAGGGGACGGA 59% conserved"
repeat_region 10745..10756
repeat_region /notes="2.0 copies 6 mer GTGGG 24% conserved"
repeat_region 10889..10903
repeat_region /notes="2.1 copies 7 mer CCGGCC 30% conserved"
repeat_region 11166..11193
repeat_region /notes="3.5 copies 8 mer TGTGTGCA 33% conserved"
repeat_region 11177..11211
repeat_region /notes="5.8 copies 6 mer GTGCAT 38% conserved"
repeat_region 11250..11261
repeat_region /notes="2.0 copies 6 mer ATGGGT 24% conserved"
repeat_region 11273..11282
repeat_region /notes="2.5 copies 4 mer TGCC 20% conserved"
repeat_region 11293..11309
repeat_region /notes="2.8 copies 6 mer TGTGTA 25% conserved"
repeat_region 11311..11411
repeat_region /notes="2.5 copies 40 mer
repeat_region TGCATGGGTGTCATGGTATGGTGTGGTGTGGTGTGCCT 202% conserved"
11323..11334
```

```
repeat_region /note="2.0 copies 6 mer ATGGGT 24% conserved"
11346..11355
repeat_region /note="2.5 copies 4 mer TGCC 20% conserved"
11368..11379
repeat_region /note="2.0 copies 6 mer ATGGGT 24% conserved"
11391..11400
repeat_region /note="2.5 copies 4 mer TGCC 20% conserved"
11431..11683
repeat_region /note="7.0 copies 36 mer
repeat_region TGCATGGTATGGGTGTGTATGTGATGGTGGTGGTG 108% conserved"
11451..11490
repeat_region /note="2.9 copies 14 mer TGTGCATGGGTGCG 71% conserved"
11452..11477
repeat_region /note="3.2 copies 8 mer GTGCATGG 29% conserved"
11485..11498
repeat_region /note="2.3 copies 6 mer TGGGTG 28% conserved"
11513..11523
repeat_region /note="2.8 copies 4 mer TGCC 22% conserved"
11521
misc_feature
11531..11562
repeat_region /note="Tandem repeat. Assembly not confirmed by digest"
11536..11635
repeat_region /note="5.3 copies 6 mer TGTGCA 32% conserved"
11595..11604
repeat_region /note="6.2 copies 16 mer GTGTGCATGTGTATGG 82% conserved"
11679..11690
repeat_region /note="5.0 copies 2 mer TG 20% conserved"

Query Match 47.1%; Score 427; DB 9; Length 101270;
Best Local Similarity 100.0%; Pred. No. 1.7e-68;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 CAGCGCCATCCACCTGTGCAACACACCCCGTCCAGAAGTACCTGCAAGATGATGGCCG 539
DB 44821 CAGCGCCATCCACCTGTGCAACACACCCCGTCCAGAAGTACCTGCAAGATGATGGCCG 44762

QY 540 CAGCCCCCTGTGCGCGCACACAAATGTGGACAGCACAGGTTCAGGAGTACCTGCA 599
DB 44761 CAGCCCCCTGTGCGCGCACACAAATGTGGACAGCACAGGTTCAGGAGTACCTGCA 44702

QY 600 GCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGCTCATCTACCCGTCCATGAAGGCCAT 659
DB 44701 GCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGCTCATCTACCCGTCCATGAAGGCCAT 44642

QY 660 CCGCCACGCCATGAAGGTGGCCCGCAGGACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCT 719
DB 44641 CCGCCACGCCATGAAGGTGGCCCGCAGGACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCT 44582

QY 720 CTACGGGGTGACTTCGTCTCTTGGGAGGGACTTCAGGCCCTGGCTGATCGAGATCAATTC 779
DB 44581 CTACGGGGTGACTTCGTCTCTTGGGAGGGACTTCAGGCCCTGGCTGATCGAGATCAATTC 44522

QY 780 CAGCCCCCAGTCACCCCGCTCCACGCGGTCCAGGCCAGCTGTGTGCACAGGTGCAGGA 839
DB 44521 CAGCCCCCAGTCACCCCGCTCCAGGCCAGCTGTGTGCACAGGTGCAGGA 44462

QY 840 GGACACCATCAAGGTGGCCCGTGGACCGCAGCTGTGTGATCGGCAATTCGAGCTCTGTG 899
DB 44461 GGACACCATCAAGGTGGCCCGTGGACCGCAGCTGTGTGATCGGCAATTCGAGCTCTGTG 44402

QY 900 GAGGCAG 906
DB 44401 GAGGCAG 44395

RESULT 3
LOCUS HSM805098 4238 bp mRNA linear PRI 12-JUL-2002
DEFINITION Homo sapiens mRNA; cDNA DKFZp586B0320 (from clone DKFZp586B0320).
ACCESSION AL833939
VERSION AL833939.1 GI:21739506
KEYWORDS
SOURCE Homo sapiens (human)
```



```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4238)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2002) GSF, Institut fuer Bioinformatik MiPS,
Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp586B0320) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source            1..4238
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp586B0320"
                     /tissue_type="uterus"
                     /clone_lib="586 (synonym: hutel). Vector pSport1; host
                     DH10B; sites NotI + SalI/MluI"
                     /dev_stage="adult"
     gene              1..4238
                     /gene="DKFZp586B0320"
     CDS                1340..2844
                     /gene="DKFZp586B0320"
                     /note="similarity to tubulin-tyrosine ligase"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD38794.1"
                     /db_xref="GI:21739507"
                     /db_xref="GOA:Q8NDN8"
                     /db_xref="UniProt/TREMBL:Q8NDN8"
                     /translation="MDIDKLEAPLYLTPEGMSLFQRYQVYVHEGAELRHLDTOVQR
                     CEDIIQQLQAVVQIDMEGRNIWIKVKGASRGRIWCMHDLBMLKLVNPNVPMK
                     DGKVVQVIERPLIFGTKPDLRQFLVTDWNPILTVMFDRDSYIRFSTQPSLKNLD
                     NSVHLNNSIQKHLENSCHRHPLLPDNNWSSORFOAHLQENGAFNASTIIVFGMKD
                     AVYHALQTSQDTVQCKRSFELYGADFVGEDFQFWLIEINASPMPASTAVARLCA
                     GVQADRLVIRDRRLNCDCAFELIYKQPAVEVPYQVIRLLVEGRTIKKPMAMCH
                     RMGVVPVAPLTLTORGSEARHFPSLHTKAQLSPHLRHQGVQLRRQ:ISKLVGTKA
                     LSTTGKALRTLPTAKVFI:SLPPNLDKVPASILKPKRVGLDL"
                     /length=4216
                     /polya_site
                     /gene="DKFZp586B0320"
ORIGIN
Query Match      44.2%; Score 400.4; DB 9; Length 4238;
Best Local Similarity 66.1%; Pred. No. 1.9e-63;
Matches 578; Conservative 0; Mismatches 296; Indels 0; Gaps 0;
QY 3 GGACATCGACAGCTGACGAGATGCGGTGGAGGACCTCACTGAGGCGGAGTGGAGACCT 62
DB 1542 GGACATCGACAGGACCTGGAGGCGCGCGTGTACTCACCCCGAGGGGTGGCTCTT 1601
QY 63 GACCCAGCAGTACTCTCCCTCGTTCTGCGGATGCTTTTCATCTCCAATTCAGAAATTA 122
DB 1602 CTTCCAGCGCTACTACCAAGTGGTCCAGAAAGGGGAGCACTCAGGACCTCGACACTCA 1661
QY 123 CTTTTCGAGTGCACGGCTCTGCTGAATAGAAATCAGCTGTGTGAACCTTCAGACGGACAT 182
DB 1662 GGTCCAGCGCTGTGAGGACATCTGCAGCAGCTGCAGGCGGTGGTATCCCGACATAGACAT 1721
QY 183 TGACGGGCTCCGGAACATCTGGAATATTAAGCCCGGCGCAAGTCCCGGGCGCCGAGACAT 242
DB 1722 GGAAGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCAAGTCCCGTGGACGAGGAT 1781
QY 243 AGTGTGCAATGGACCGTGTGGAGAGATCTTGAGAGCTGGCAGCTGCAGACCCCTCTTTC 302
DB 1782 CATGTGTCATGGACCACTTGGAGGAGATGCTGAAGTGGTGAACGGCAACCCCGTGGTGTAT 1841
```

```
QY 303 CAGGACAAACAAGTGGTGGTCCAGAAATGATCAGAGACGCGCTGCTCATCTGTGACAC 362
DB 1842 GAAGACAGCGCAAGTGGTGGTGCAGAAATATATTGAGGCGGCCCTCCTCATCTTTGGCAC 1901
QY 363 CAAGTTGCACATCAGACAGTGGTTCCTCGTACGAGACTGGAAACCCCTCAACCATCTGGTT 422
DB 1902 CAAGTTTGACCTTCAGACAGTGGTTCCTCGTAACTGACTCTGGAACCCCATTTACCGTGGTT 1961
QY 423 CTACAAGGAGAGTACTTTCGCGTCTCTCAACTCAGCGCTTCTCCCTGGAGCAAGCTGACAG 482
DB 1962 CTACCGGACAGCTATATCCGCTTTTTCACGAGCGCCCTTCTCCCTGGAAGAACCTGGACAA 2021
QY 483 CGCCATCCACCTGTGTCAACAACCGCGTCCAGAAATGATCTCTGAAAGAAATGATGTGGCGCGCAG 542
DB 2022 CTCAGTGCACCTGTGTCAACAACCTCCATCCAGAAGCACCTGGAGAACTCATGCGCATCGGCA 2081
QY 543 CCCCTGTGTCGCCGACACAAACATGTGGACGAGCAGGTTCCAGGAGTACTCTGACGCG 602
DB 2082 TCCACTGTCTTCGCCGACAAACATGTGTCTTAGCCAGAGGTTCCAGGCCACCTCGAGGA 2141
QY 603 CCAGGCGCGTGGCGCGCTGTGGGCGAGCGTCACTACCCGTCATGAAGAAGGCCATCGC 662
DB 2142 GATGGTGGCCCCAAATGTGTGTCCACCATCATCTGCTGCGCANTGAAGGATGCTGTGAT 2201
QY 663 CCACGCCATGAAGTGGGCCAGGACCACTGGAGCGCTCGCAAGAACAGCTTTGAGCTCTA 722
DB 2202 CCACGCACCTTCAGACCTCCAGGACACCGTGCAGTGTGGAAAGGCCAGCTTTGAGCTCTA 2261
QY 723 CGGGCTGACTTCGTCTTGGGAGGAGCTTCAGGCGCTGGCTGATCGAGATCAATCCAG 782
DB 2262 TGGCGCTGACTTCGTGTTCGGGAGGAGCTTCAGGCGCTGGCTGATCGAGATCAAGCCAG 2321
QY 783 CCCACCATGACACCGCTGCAGCGCTGCAGCGCCAGCTGTGTGCACAGGTGCAGGAG 842
DB 2322 CCCACGATGGACCTTCACAGCAGTCACTGCCCCGCTCTGTGTGGCTGCAAGCTGA 2381
QY 843 CACCATGAAGTGGCGCTGGAGCCGAGCTGTGAC 876
DB 2382 CACCCTGCGGTGGTCAATTGACCGGAGCTGAGC 2415
RESULT 4
LOCUS BD158727 2326 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD158727
VERSION BD158727.1 GI:27864485
KEYWORDS JP 2002191363-A/13570.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
JOURNAL Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
COMMENT Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 13570 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/13570
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000208090
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,S
PC KEIICHI NAGAI,TETSUJI OTSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10,
PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
```



```
FEATURES
  source
    FT CDS Location/Qualifiers
        1..2326
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
ORIGIN
Query Match 44.0%; Score 398.8; DB 6; Length 2326;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY 3 GGACATCGACACGTACGACATCCGTTGGAGGACCTCCTCAGCGCCGAGTGGAGGACCT 62
DB 732 GGACATCGACACGACCTCGAGGCGCCGCTGTACCTCACCCTCGAGGCTGGTCCCTCTT 791
QY 63 GACCCAGCAGTACTACTCCCTCGTTCATGCGCATGCTTTTCATCTCCAATTCAAGAAATTA 122
DB 792 CCTCCAGCGCTACTACCAAGTGTCCACCAAGGGGAGAACTCAGGCACCTCGACACTCA 851
QY 123 CTTTTCGAGTGCAGGCTCTCTCTGAATAGATCAGCTCTGTGAACCTCAGACGACAT 182
DB 852 GGTCCAGCGCTGTGAGGACATCTCTGACGAGCTGCGAGGCGGTGGTACCCAGATAGACAT 911
QY 183 TGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGCAAGTCCCGGGCCGAGACAT 242
DB 912 GGAAGGGGATCGCAACATCTGGATCGTGAAGCAGAGCCAAAGTCCCGGGAGCGGACAT 971
QY 243 AGTGTGCATGAGCGGTGTGGAGGAGATCTGGAGCTGGCAGCTCGAGACACCCCTCTTTC 302
DB 972 CATGTGCATGAGCACTCCTGGAGGAGATGCTGAAGTGGTGAACGGCAACCCCGTGTGAT 1031
QY 303 CAGGAGCAACAAAGTGGGTGGTCCAGAAATGATCAGAGACCGCGCTGCTCATCTGTGACAC 362
DB 1032 GAAGGACGGCAAGTGGGTGGTGCAGAAATATTTGAGCGGCCCTCTCTCATCTTTGGCAC 1091
QY 363 CAAGTTTCGATCAGACAGTGGTTCCTCGTCAAGTGGTGAACCCCTGACCATCTGTT 422
DB 1092 CAAGTTTGACCTCAGACAGTGGTTCCTGGTAACTGATGGAACCCACTTACCGTGGTT 1151
QY 423 CTACAAGGAGAGTTACTTGGCGTCTCAACTCAGCGCTTCTCCCTCGACCAAGCTGACAG 482
DB 1152 CTACCGGACAGTATATCGCTTTTCCAGCGAGCCCTTCTCCTGAGAACCTGGACAA 1211
QY 483 CGCCATCACTGTGCAACAAACCGCTTCCAGAAATGATGAGAAATGATGTTGGCGCGCAG 542
DB 1212 CTCAAGTGCACCTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCCATCGCA 1271
QY 543 CCCCCTGCTGCCGACACAAACATGTGGACCAACAGTTCAGGAGTTCAGGAGTACCTGACGG 602
DB 1272 TCACCTGCTTCCGCCAGACAAACATGTGGTCTAGCCAGAGGTTCCAGGCCACCTGCAGGA 1331
QY 603 CCAGGCGCTGGCGCGTGTGGGGGACGCTCATCTACCCGCTCCATGAAGAAGCCATCGC 662
DB 1332 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCGTGGCTGGATGAGATGCTGTGAT 1391
QY 663 CACGCGCATGAAGGTGGCCAGGACCAACGCTGGAGCCTTCGAGAAACAGCTTTGAGCTCTA 722
DB 1392 CCAAGCACTTCAGACCTCCAGGACACCGTGCAGTGTCCGAGAGCCAGCTTTGAGCTCTA 1451
QY 723 CGGGGCTGACTTCTGCTTGGGAGGACCTTCAGGCCCTGGCTGATCGAGATCAATCCAG 782
DB 1452 TGCGCTGACTTCTGCTTGGGAGGACCTTCAGGCCCTGGCTGATCGAGATCAATCCAG 1511
QY 783 CCCACCATGACCCCGTCCAGCGCGTTCAGGCCCTGAGCTGTGTGACAGCTGTCAGGAGGA 842
DB 1512 CCCCAGATGGACACCTTCACAGCACTCACTGCCCCGCTCTGTGCTGGCTGCAAGCTGA 1571
QY 843 CACCATCAAGTGGCGGTGGACCGCA 868
DB 1572 CACCTCGCGTGGTCAATTGACCGGA 1597
```

```
RESULT 5
AX881015
LOCUS
DEFINITION
Sequence 15920 from Patent EP1074617.
ACCESSION
AX881015
VERSION
AX881015.1 GI:40035751
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
1
AUTHORS
Ota, T., Isegai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, I.
TITLE
Primers for synthesizing full-length cDNA and their use
JOURNAL
Patent: EP 1074617-A 15920 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..2326
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
94..>2325
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE92353.1"
/db_xref="GI:40035752"
/translation="MRLNRAKIYVERAVKQKIFTIOGCPVIRCLRRRGWVEKMH
VHRSGTLLPQKDLSSAMGSDTDEDEDEFOPSQLPFDPLDLKFDLDDGTH
ALMVLGLNRLNLPWFEDVANSFPRCYCLGAEEDDKAFIDFWLTAAARNVLKLVK
SWKSPYQAVVEEASGQPKQKQKPNVLSFVDFDEALCAEEVLSNLAHMDIDKD
LSAPLTPGWSLFLQYVHGAELRHLDTVQRCEDILQQVAVPQIDMSGD
RNWIVKPAKSRGIMCMHLEMLKLVNPNVVMKDGKRWVYKNIQKLENSHR
FLPDLQFWLTDWNPFTVYRDSYIRFTQFSLKNDLVHLCNNISQDTCQKRSF
HPLPDNNWSSORFOAHLEQAPNAWSTIIVPGMKDAVIALQTSQDTCQKRSF
ELYGADVFGEDEFPWLLEINASTPMASTAVTARLCAGVQADTLRVLDRLDNCD
TCAPHLIYQPAVEVPQYVIRLLVEGFTIKKPMAMCHRRMGRVPAVPLLTQSGEA
RHFFSLHTKAQLPSFHLRHQGVLRHSLKLVTKALSTTGKALURTLPTKAFVLSL
PNPLDFKVAFLKPRKAPALCLRGQLEVPFCLPKESEQFLAPVGRSPKANRSP
DCDKPRAEACPMKRLSLPLPLVGTGTRRRGLGDMKLGKPLLRLLPFTALVLDPTPNKK
KQVK"
ORIGIN
Query Match 44.0%; Score 398.8; DB 6; Length 2326;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY 3 GGACATCGACACGTACGACATCCGTTGGAGGACCTCCTCAGCGCCGAGTGGAGGACCT 62
DB 732 GGACATCGACACGAGACCTGGAGGCGCCGCTGTACCTCACCCTCGAGGCTGGTCCCTCTT 791
QY 63 GACCCAGCAGTACTACTCCCTCGTTCATGCGCATGCTTTTCATCTCCAATTCAAGAAATTA 122
DB 792 CCTCCAGCGCTACTACCAAGTGTCCACCAAGGGGAGAACTCAGGCACCTCGACACTCA 851
QY 123 CTTTTCGAGTGCAGGCTCTCTCTGAATAGATCAGCTCTGTGAACCTCAGACGACAT 182
DB 852 GGTCCAGCGCTGTGAGGACATCTCTGACGAGCTGCGAGGCGGTGGTACCCAGATAGACAT 911
QY 183 TGACGGGCTCCGGAACATCTGGATTATTAAGCCCGCGCAAGTCCCGGGCCGAGACAT 242
DB 912 GGAAGGGGATCGCAACATCTGGATCGTGAAGCAGAGCCAAAGTCCCGGGAGCGGACAT 971
QY 243 AGTGTGCATGAGCGGTGTGGAGGAGATCTGGAGCTGGCAGCTCGAGACACCCCTCTTTC 302
DB 972 CATGTGCATGAGCACTCCTGGAGGAGATGCTGAAGTGGTGAACGGCAACCCCGTGTGAT 1031
QY 303 CAGGAGCAACAAAGTGGGTGGTCCAGAAATGATCAGAGACCGCGCTGCTCATCTGTGACAC 362
DB 1032 GAAGGACGGCAAGTGGGTGGTGCAGAAATATTTGAGCGGCCCTCTCTCATCTTTGGCAC 1091
QY 363 CAAGTTTCGATCAGACAGTGGTTCCTCGTCAAGTGGTGAACCCCTGACCATCTGTT 422
```


Db 1092 CAAGTTTGACCTCAGACAGTGGTTCTCTGGTAACCTGACCTGGAACCCACTTACCGTGTGGTT 1151
QY 423 CTAAAGGAGAGTACTTGGGGTCTCTCAACTCAGCGCTCTTCCCTCGAACAAGCTGACAG 482
Db 1152 CTACCGGACAGCTATATCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACTGACAA 1211
QY 483 CGGCATCCACTGTGTGCAACAAACCGCGTCCAGAAAGTACCTGAAGAATGATGTGGCGCGCAG 542
Db 1212 CTCAGTGCACCTGTGCAACAACTCCATCCAGAAAGCACTCGGAGAACTCATGCGCATCGCA 1271
QY 543 CCCCTGCTGCCCGCACACAAACATGTCGACCAAGTTCAGAGTTCAGAGTACTCTGCGAGCG 602
Db 1272 TCACACTGCTTCCGCCAGAACATATGGTGTCTAGCCAGAGGTTCAGGCCACCTCGCAGGA 1331
QY 603 CCAGGCGCGTGGCGCGCTGTGGGCGACGCTCATCTACCGTCCATCAAGAAGCCATCGC 662
Db 1332 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCTGCTGCTCGCATGAATGCTGTGAT 1391
QY 663 CCACGCCATGAAGGTGGCCAGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTA 722
Db 1392 CCACGCACCTTCAGACCTCCAGGACACCGCTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTA 1451
QY 723 CGGGGTGACTTCTGCTTCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAG 782
Db 1452 TGGCGTGTACTTCTGCTTCTGCGGAGGACTTCCAGCCCTGGCTGATTCAGATCAAGCCCA 1511
QY 783 CCCCACATGACCCCTGCACGCGGTTCACGCGCCAGCTGTGTGACAGGTGCAGGAGGA 842
Db 1512 CCCCACATGACCCCTGCACGAGTCACTGCGCGGCTGTGTGCTGGGTGCAGACTGA 1571
QY 843 CACCATCAAGGTGGCGGTGGACCGCA 868
Db 1572 CACCCTGCGGTGGTCAATGACCGGA 1597

RESULT 6
AK023960
LOCUS Homo sapiens cDNA FLJ13898 fis, clone THYRO1001738, weakly similar
DEFINITION to TUBULIN--TYROSINE LIGASE (EC 6.3.2.25).
ACCESSION AK023960
VERSION AK023960.1 GI:10436083
KEYWORDS oligo cappings; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratani, A., Sudo, H., Hosoiri, T., Kaku, Y., Yokoi, T., Furuya, T.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togliya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, S.,
Shiohata, N., Sano, S., Moriya, S., Momiya, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigetate, K., Senba, T., Matsuura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,

Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
2
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,
Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K.,
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Oshima, A.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2326)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry for Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
LOCATION/Qualifiers
1. .2326
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYRO1001738"
/tissue_type="thyroid gland"
/clone_lib="THYRO1"
/note="cloning vector: pME18SFL3"
94. .>2326
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14741.1"
/db_xref="GI:10436084"
/translation="MNRLNAKIYVERAVKQKFIITQCVPIRLRRRGWVEKRM
VHRSGLTLLPKOKLDSAMGSDTDEDEDEEFQPSQDFDLDLLKFDLDLGT
ALMVGLCLNRLNLPWFDEVDAMSPPRCVCLGAEDDKAFIDFWLTAARNVLKLVK
SEWKSYPICQAVEEASGDKPKQKSNPVLVSPFVDRALCALCEVLSNLAHMDIKD
LEAPYLTPGEWSLFLQYQVWHEGAEHLDTQVRCEDLIQQLQAVVPDIDEGD
RNMIWPKGASRGSGIMCMOHEMLKLVNPNVMDKRWVQKITYERPLLIPTGK
FDLRQWFLVTWNPLTVWFYRDSYTRFSTQPSLKNLNSVHLNNSIQHLENSCH
HPLPPDNWSSQRFQHLQMGAPNAWSTIIPGMDAVIHALQTSQTVQCRKASF
ELYGADPVGEDFQPLWIEINASTPMAPSTAVTLACAGVQADTLRVVLDRLNCD
TGAFELIVQPAVEVPOVGIIRLVVEGFTIKKPMACHRMGMVRPAVLLTORGSEA
RHHFSLTKAQLPSFHLVHQGVIRRHQSKLVTKALSTGKALRILPTAKYFVLSL
PPNLDKVPASILKPRKPAULLGRPOLEVPCCCLPKSEQFLAPVGRSRPANSRP
DCDKPRABACPMKRLSPLKPLFLVGTQRRRGLGDMKGLKPLLRPTALVLDPTPNKK
QVK"

ORIGIN
Query Match 44.0%; Score 398.8; DB 9; Length 2326;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
QY 3 GGACATCGACACGTCAGCAGATGCGCTGGAGGACCTCACTGAGGCGGAGTGAGGACCT 62
Db 732 GGACATCGACAGGACCTTGAGGCGCCCGCTGTACTCACCCTGAGGCTGTCTCTT 791
QY 63 GACCCAGCAGTACTACTCCCTCGTTTCATGCGGATCGTTTCATCTCAATTCAGAAATTA 122
Db 792 CCTCCAGCGCTACTACCAAGTGGTCCACGAGGGGAGAGCACTCAGCACCTCGACTCA 851
QY 123 CTTTTCGCGATGCCAGGCTCTCTCTGAATAGATCAGTCTGTGAACCTTCAGACGACAT 182

Db 852 GGTCCAGCGCTGTGAGGACATCTCTGACAGAGCTGCGAGCGCGTGTATCCCCAGATAGACAT 911
Qy 183 TGACGGGCTCGGAAATCTGGATTATTAAGCCCGCGGCCCAAGTCCCGGGCCGAGACAT 242
Db 912 GGAAGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCCCAAGTCCCGCGGAGCGAGCAT 971
Qy 243 AGTGTGCATGAGACCGTGTGGAGGAGATCTTGGAGCTGCGAGCTGCGACACACCTCTTTTC 302
Db 972 CATGTGATGACCACTTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGTAT 1031
Qy 303 CAGGGAACAAGTGGGTGGTCCAGAAAGTATCATCGAGACCGCGTGTCTCATCTGTGACAC 362
Db 1032 GAAGGACGGCAAGTGGGTGGTGCAGAAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCAC 1091
Qy 363 CAAGTTCGACATCAGACAGTGGTTCCTGTCGACGGAATGGAACCCCCCTGACCATCTGGTT 422
Db 1092 CAAGTTTGAACCTCAGACAGTGGTTCCTGTTAACTGACTGGAACCCACTTACCCGTTGGTT 1151
Qy 423 CTACAAAGGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGACAGAGCTGGACAG 482
Db 1152 CTACCGGACAGCTATATCCGCTTTTCCAGCAGCCCTTCTCCCTGGAAGAACTTGGACAA 1211
Qy 483 CGCCATCCACTGTGCAACACCGCGTCCAGAAAGTACTGGAAGTACTGGAAGTACTGAGGCGCCAG 542
Db 1212 CTAGTGCACCTGTGCAACAACTCCATCCAGAGCACTCGAGAACTCATGCCATCGCA 1271
Qy 543 CCCCTGCTGCCGCAACAACATGTTGGAACAGCAGGTTCCAGAGTACTCTGACAGCG 602
Db 1272 TCCACTGCTTCCGACAGCAACATGTGTCTAGCAGAGGTTCCAGGCCACCTGACGGA 1331
Qy 603 CAGGGCGGTGGCGCGTGTGGGCGAGGTCATCTACCCGTCATGGAAGAGGCCATCGC 662
Db 1332 GATGGGTGCCCCAAATCTTGTGTCACATCATCTGTCCTGGCATGAAGATGCTGTGAT 1391
Qy 663 CCAGCCATGAAGTGGCGCCAGGACCACTGAGGCGCTCGAAGAAACAGCTTTGAGCTCTA 722
Db 1392 CCAGCACTTCAGACCTCCAGGACACGTCAGTGTGGAAGCCAGCTTTGAGCTCTA 1451
Qy 723 CGGGCTGACTTGTCTTGGGAGGACTTTCAGGCGCTGCTGATCAGATCAATTCAG 782
Db 1452 TGGCGCTGACTTGTGTGCGGGAGGACTTCCAGCCCTGGCTGATTCAGATCAACGCCAG 1511
Qy 783 CCCACCATGACCCGTCACCGCGTCCAGGCGCTGTCAGCCAGCTGTGTCACAGGTCAGGAGGA 842
Db 1512 CCCACCATGACCCCTCCACAGCAGTCACTGCCCGCTCTGTGCTGGCGTGAAGCTGA 1571
Qy 843 CACCATCAAGTGGCGCTGGAACGCCA 868
Db 1572 CACCTGGCGTGTGATTCAGCCGA 1597

RESULT 7
AX301197 2380 bp DNA linear PAT 30-NOV-2001
LOCUS Sequence 39 from Patent W00185942.
DEFINITION AX301197
ACCESSION AX301197
VERSION AX301197.1 GI:17382288
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Yue,H., Tang,Y.T., Au-Young,J., Lu,D.A., Baughn,M.R., Hillman,J.L.,
Azimzai,Y., Lal,P., Yao,M.G., Bandman,O., Burford,N., Batra,S.,
Kearney,L. and Policky,J.L.
Cytoskeleton-associated proteins
Patent: WO 0185942-A 39 15-NOV-2001;
Incyte Genomics, Inc. (US)
Location/Qualifiers
1. .2380
/organism="Homo sapiens"
/mol_type="unassigned DNA"

ORIGIN
Query Match 44.0%; Score 398.8; DB 6; Length 2380;
Best Local Similarity 66.3%; Pred. No. 3.9e-63;
Matches 574; Conservative 0; Mismatches 292; Indels 0; Gaps 0;
/db_xref="taxon:9606"
/note="Incyte ID No: 2156553CB1"
Qy 3 GGCATCGACACGTGACGATGCGTGGAGGACCTCCTCAGTGGCCGAGTGGGAGGACCT 62
Db 473 GGCATCGACACGAGGACCTGGAGGCCCGCTGTACCTCACCCCGAGGCGTGGTCCCTCTT 532
Qy 63 GACCCAGCAGTACTACTCCCTCGTTCATGCGATGCTTTCATCTCCAAATTCAGAAATTA 122
Db 533 CTTCCAGCGCTACTTACCAAGTGTCCAGAAAGGGGAGAACTCAGGCACTCTGCACTCA 592
Qy 123 CTTTTCGAGTGCACGCTCTGCTGAATCAAGTCAAGTCTGTGAACCCCTCAGACGACAT 182
Db 593 GGTCCAGCGCTGTGAGGACATCTCGCAGCAGCTGCAGGCCGTGGTACCCAGATAGACAT 652
Qy 183 TGAAGGCTCCGAAACATCTTGGATTATAAAGCCCGCGGCCAAGTCCCGGGCCGAGACAT 242
Db 653 GGAAGGGGATCGCAACATCTGGATCGTGAAGCAGGAGCCAAAGTCCCGTGGACGAGCAT 712
Qy 243 AGTGTGCATGACCGCTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCACTCTTTC 302
Db 713 CATGTGATGGAACCACTGTGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGT 772
Qy 303 CAGGGAACAAGTGGGTGTCCAGAAAGTACATCGAGACCGCGCTGTCTCATCTGTGACAC 362
Db 773 GAAGGACGGCAAGTGGGTGTGAGAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCAC 832
Qy 363 CAAGTTCGACATCAGACAGTGGTTCCTGTCACGAGCTGGAAACCCCTGACCATCTGGTT 422
Db 833 CAAGTTTGACCTCAGACAGTGGTTCCTGTTAACTGACTGGAACCCCACTTACCCGTGGT 892
Qy 423 CTACAAAGGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGACAAAGCTGGACAG 482
Db 893 CTACCGGACAGCTATATCCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAACCTTGAACA 952
Qy 483 CGCCATCCACTGTGCAACAAACCGCTCCAGAAAGTACTCGAAGAAATGATGTGGCGCCGAG 542
Db 953 CTCAGTGCACCTGTGCAACAACTCCATCCAGAGCACCTGGAGAACTCATGCCATCGCA 1012
Qy 543 CCCCTGTGCTCCGCAACAACATGTGGACCAAGTTCAGAGTTCAGAGTACCTGACAGCG 602
Db 1013 TCCACTGTCTCCGCAACAACATGTGGTCTAGCAGAGGTTCCAGGCCCACTGACGGA 1072
Qy 603 CAGGGCGGTGGCGCGTGTGGGCGAGCGTCACTACCCGTCATGAAGAGGCCATCGC 662
Db 1073 GATGGGTGCCCCAAATGCTTGGTCCACCATCATCTGCTGGCATGAAGATGCTGTGAT 1132
Qy 663 CCACGCGCATGAAGGTGGCCCGCAGGACCAAGTGGAGCCCTCGCAAGAAACAGCTTTGAGCTCTA 722
Db 1133 CCACGCACTTCAGACCTCCCGACAGACCGTGTGAGTGTGGAAGGCCAGCTTTGAGCTCTA 1192
Qy 723 CGGGGCTGACTTCTGCTTGGGAGGAGTTCAGGCCCTGCTGATCAGATCAATTCAG 782
Db 1193 TGGCGCTGACTTCTGTTTGGGAGGACTTCCAGGCCCTGGCTGATTCAGATCAACGCCAG 1252
Qy 783 CCCACCATGACCCCTGCAAGCGGCTCAGCGCCAGCTGTGTCACAGGTCGAGGAGGA 842
Db 1253 CCCACCATGAGCACCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGAAGCTGA 1312
Qy 843 CACCATCAAGTGGCGCTGGACCGCA 868
Db 1313 CACCTGGCGGTGCTCATTCAGCCGA 1338

RESULT 8
HSM800637 2553 bp mRNA linear PRI 18-FEB-2000
LOCUS HSM800637
DEFINITION Homo sapiens mRNA; cDNA DKFZp434B103 (from clone DKFZp434B103);

ORIGIN		Query Match		44.0%; Score 398.8; DB 6; Length 2848;		Best Local Similarity 66.0%; Pred. No. 3.8e-63;		Matches 577; Conservative 0; Mismatches 297; Indels 0; Gaps 0;	
QY	3	GGACATCGACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGCCGAGTGGGAGGACCT	62						
Db	1353	GGACATCGACACGACCTGAGGCGCCGCTGTATCTCACCCCGAGGCGTGGCCCTCTT	1412						
QY	63	GACCCAGCAGTACTACTCCCTCGTTTCATGCGGATGCTTTCATCTCCAAATTCAGAAATTA	122						
Db	1413	CCTCCAGCGCTACTACCAAGTGTCTCAGAGGGGCGAGACTCAGGCACCTCGACACTCA	1472						
QY	123	CTTTTCGAGTCCGACGCTCTGCTGAATAGATACTCTGTGTGAACCTCTCAGAGGACAT	182						
Db	1473	GGTCCAGCGCTGTGAGACATCTCTGCAGCAGCTGTCAGCGCGCTGTATCCCGAGATAGACAT	1532						
QY	183	TGACGGGCTCCGGAACATCTGGATTATAAGCCGCGGCCCAAGTCCCGGGGCGGACAT	242						
Db	1533	GGAAGGGGATCGCAACATCTGATCTGTGAAGCCAGGACCAAGTCCCGTGGACGAGGCAT	1592						
QY	243	AGTGTGATGGACCGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCACACCACTCTTTTC	302						
Db	1593	TATGTGATGGACCACTTGAGGAGATGCTGAAGCTGTGNAACGGCAACCCCGTGTGAT	1652						
QY	303	CAGGACAAACAGTGGTGTCCAGAAAGTACATCGAGACCGCGCTGCTCATCTGTGACAC	362						
Db	1653	GAAGGACGCGCAAGTGGTGTGTCAGAAAGTATATTGAGCGGCCCTCTCATCTTTGGCAC	1712						
QY	363	CAAGTTTCATCGACAGTGGTTCCTCGTCAAGGACTGGAACCCCTGACCATCTGGTT	422						
Db	1713	CAAGTTTTCATCGACAGTGGTTCCTCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT	1772						
QY	423	CTACAGGAGGAGTTACTTTCGCTTCTCAACTCAGCGCTTCTCCTGTCAGCAAGCTGGACAG	482						
Db	1773	CTACCGGACAGCTATATCGCTTTTCCAGCGAGCCCTTCTCCTGAAGAACTGGACAA	1832						
QY	483	CGCCATCCACTGTGCAACACCGCGTCCAGAAAGTACTCTGAAGAAATGATGGGCGCGCAG	542						
Db	1833	CTCAGTGCACCTGTGCAACAACTCCATCCAGAGACCTCGGAGAACTCATGCGCATCGCA	1892						
QY	543	CCCCCTGCTCCCGCACACATCTGGAGCCAGCACAGGTTCCAGAGTACCTGACGG	602						
Db	1893	TCCACTGCTTCCGCGACACAACTATGTGTAGCTAGCCAGAGTTCCAGGCCACCTGCGAGGA	1952						
QY	603	CCAGGCGCGTGGCGCGTGTGGGCGCGTCTACCTACCGTCCATGAAGAAGCCCATCGC	662						
Db	1953	GATGGGTGCCCCAAATGCTTGTGTCACCATCATGTGCTGCGCATGAGGATGCTGTGAT	2012						
QY	663	CCACGCGCATGAAGTGGCCCGACGACCACTGGAGCCCTCGCAAGAACACAGCTTTGAGCTCTA	722						
Db	2013	CCACGCACTTCAGACCTCCCGAGGACCGTGCATGTCGAGAGCCAGCTTTGAGCTCTA	2072						
QY	723	CGGGCTGACTTGTCTTGGAGGAGACTTCAGGCCCTGCTGATCGAGATCAATTCAG	782						
Db	2073	TGGCGCTGACTTGTGTGTCGGGGTGGACTTCCAGCCCTGCTGATTTGAGATCAACGCGCAG	2132						
QY	783	CCCACCATGACCCCGTCCAGCGCGTCCAGGCCCGCTGTGTGCAAGTGTGAGGAGGA	842						
Db	2133	CCCACGATGGACCTTCCAGAGCAGTTCATGCCCGGCTCTGTGCTGGCGTGCAGAGCTGA	2192						
QY	843	CACCATCAAGGTGGCGCTGGACCGCAGCTGTGAC	876						
Db	2193	CACCTGCGGTGTCTATTGACCGGAGCTGGAC	2226						

RESULT 10
AK097236
LOCUS
DEFINITION
Homo sapiens cDNA FLJ39917 fls, clone SPLEN2019405, highly similar
to Homo sapiens HOTT1 protein mRNA.
ACCESSION
AK097236

VERSION
KEYWORDS
SOURCE
ORGANISM

AK097236.1 GI:21756925

oligo capping; fis (full insert sequence).

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hiraoka, K., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musaashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Moniyama, H., Satoh, N., Takami, S.,
Teraehima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)

TITLE

JOURNAL
NAT. GENET. 36 (1), 40-45 (2004)

REFERENCE
AUTHORS

Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,
Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A.,
Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K.,
Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y.,
Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished

TITLE
REFERENCE

3 (bases 1 to 2848)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES
Location/Qualifiers

1..2848
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SPLEN2019405"
/tissue_type="spleen"

AR339003	LOCUS	AR339003	3001 bp	DNA	linear	PAT 17-AUG-2003
	DEFINITION	Sequence 494 from patent US 6569662.				


```
Db 2143 CCCCCAGTGGCACCCTCCACAGCAGTCACTGCCCGGCTGTGCTGGCTGCAAGCTGA 2202
Qy 843 CACATCAAGTGGCCCTGGACGCCA 868
Db 2203 CACCTCGCGGTGGTCAATTGACCGGA 2228

RESULT 12
AF078842 1684 bp mRNA linear PRI 09-JAN-2000
DEFINITION Homo sapiens HOTT1 protein mRNA, complete cds.
ACCESSION AF078842
VERSION AF078842.1 GI:6683744
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1684)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
YU, Y., ZHANG, C., LUO, L., OUYANG, S., ZHANG, S., LI, W., WU, J.,
ZHOU, S., LIU, M. and HE, F.
TITLE Functional prediction of the coding sequences of 50 new genes
deduced by analysis of cDNA clones from human fetal liver
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1684)
AUTHORS YU, Y., ZHANG, C., LUO, L., OUYANG, S., ZHANG, S., LI, W., WU, J.,
ZHOU, S., LIU, M. and HE, F.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-1998) Dept. of Experimental Hematology, Beijing
Institute of Radiation Medicine, 27 Taiping Rd, Beijing 100850,
P.R.China
FEATURES
source Location/Qualifiers
1..1684
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HQ207"
/tissue_type="liver"
/dev_stage="fetus"
1..398
399..1277
/notes="hdp0207; similar to bovine and pig tubulin-tyrosine
ligase (TTL): Swiss-Prot Accession Numbers P38584 and
P38160"
/evidence="not experimental"
/product="HOTT1 protein"
/protein_id="AAF23353.1"
/db_xref="GI:6683745"
/translators="MEGDRNFWIKPGAKSRGRIMCQDHLBMLKLVNPNVVMKDG
KVVQYIERPLIFGKFLQWFLVTDWNLTVFVYRDSYIRFQPSFLKLDNS
VHLNNSIQHLSCHRLPLPDDNNSSORFOAHLQEMGAPNASTIIVPGMDAV
IHALOTSODVQCRKASFEYLGADVFQDLPWLEINASTPASTAVTARLCAGV
OADTLRVLDRLDNCDTGAFELIYKQVPTTSPASTPRSCLLPMSYDTRASSDDS
TASWALRPCRQARP"
1278..1684

3'UTR
ORIGIN
Query Match 41.6%; Score 376.8; DB 9; Length 1684;
Best Local Similarity 69.3%; Pred. No. 4.3e-59;
Matches 513; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

Qy 129 GCAGTGCAGGCTCTGCTGAATAGAATCAGCTCTGTGAACCTCCAGACGGACATTGACGG 188
Db 347 GCGCTGTGAGGACATCTCGCAGCAGCTGCAGGCGGTGTACCCAGATAGACATGGAAGG 406
Qy 189 GCTCCGGAACATCTGGATTATTAAGCCCGCCGCAAGTCCCGGGCCGAGACATAGTGTG 248
Db 407 GGATCGCAACTCTTGGATCTGTGAAGCCAGGACCAAGTCCCGGGGAGGACGATCATGTG 466
Qy 249 CATGGACCGTGTGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCCCTCTTCCAGGGA 308
```

```
Db 467 CATGGACCACCTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGAAGGA 526
Qy 309 CAACAAGTGGTGGTCCAGAGTACATCGAGAGCGCGTCTCATCTGTGACACCAAGTT 368
Db 527 CGGCAAGTGGTGGTGGCAGAAATATATTGAGCGCGCCCTCTCATCTTTGGCACCAGTT 586
Qy 369 CGACATCAGACAGTGGTTCTCGTCACGGACTGGAAACCCCTCAGACCATCTGGTTCTACAA 428
Db 587 TGACCTCAGACAGTGGTTCTCGTAACTGACTGGAACCCACTTACCGTGGTCTTACCG 646
Qy 429 GGAGAGTTACTTCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGACAGCGCAT 488
Db 647 CGACAGCTATATCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCTGACAACTCAGT 706
Qy 489 CCACCTGTGCAACAACGCCGCTCCAGAGTACCTGGAAGAATGATGTGGGCGCAGCCCT 548
Db 707 GCACCTGTGCAACAACCTCCATCCAGAAACCTGGAGAACTCATGCGCATCCCACT 766
Qy 549 GCTGCCCGCACACAACATGTGGACAGCACAGCTTCCAGGAGTACCTGAGCGCCAGGG 608
Db 767 GCTTCCCGCAGACACATATGTTGTTAGCCAGAGTTCCAGGCCACCTGAGGAGATGGG 826
Qy 609 CCGTGGCGCGTGTGGGCGAGCGCTCATCTACCCGCTCCATGAAGAAGCCCATCGCCACGC 668
Db 827 TGCCCCAATGCTTGTGCCACCATCATCGTGCCTGGCATGAAGATGCTGTGATCCACGC 886
Qy 669 CATGAAGTGGCCCGCAGGACCACTGAGCGCTCCGAGAGACAGCTTTGAGCTTACGGGG 728
Db 887 ACTTCAGACCTCCCGCAGGACACCGCTGCAAGTGTGCGAAGGCCAGCTTTGAGCTCTATGGCG 946
Qy 729 TGACTTCTGCTTGGGAGGAGCTTTCAGGCCCTCGCTGATCGAGATCAATTCAGCCCCAC 788
Db 947 TGACTTCTGCTTGGGAGGAGCTTTCAGGCCCTCGCTGATCGAGATCAATTCAGCCCCAC 1006
Qy 789 CATGCAACCCGTCACCGCGTCAACGCCCGCTGTGTGCAAGTGTGAGGAGGACACCAT 848
Db 1007 GATGGCACCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGGTGCGAAGCTGACACCT 1066
Qy 849 CAAGTGGCGCGTGGACCGCA 868
Db 1067 GCGCGTGGTCAATTGACCGGA 1086

RESULT 13
BC006830
LOCUS
DEFINITION Mus musculus RIKEN cDNA 483341J24 gene, mRNA (cdna clone MGC:11856
IMAGE:3597662), complete cds.
ACCESSION BC006830
VERSION BC006830.1 GI:13905089
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1897)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Scapleton, M., Soares, M.B., Donald, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Tavakoli, S.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., Loquellano, N.A., Peters, G.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S.D., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.C., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
```


TITLE
Schnersch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE
22388257
REFERENCE
12477932
AUTHORS
2 (bases 1 to 1897)
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (27-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 16 Row: G Column: 10.

FEATURES

Location/Qualifiers
1..1897
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:11856 IMAGE:3597662"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal Carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..1897
/gene="4833441J24Rik"
/db_xref="LocusID:101100"
/db_xref="MGI:2141418"
590..1390
/gene="4833441J24Rik"
/codon_start=1
/product="4833441J24Rik protein"
/protein_id="AAH06830.1"
/db_xref="GI:13905090"
/db_xref="LocusID:101100"
/db_xref="MGI:2141418"
/translation="MEGDQNIWIKVPGAKSRGIMCMNRLDEMLKLVDCNPMKMDG
KVIQVYIERPLIFGTFKDLQWPLDWNPLTVWFYRDSVIRFSTOPFSLKNDNS
VHLNNSIORHLEASHPEMLPDNMWSSQFOAHLQEVDAKAWSSVIVPMKAAY
IHALQTSQNVQCRASFELYGADFVFGEDFQFLIEINASTMAPSTAVTARLCAGV
QADTLURVDIRLDRSCDGFAGELIYKVRCSSTSHKQHPGPGWCET"
590..1270
/gene="4833441J24Rik"
/note="TTL; Region: Tubulin-tyrosine ligase family.
Tubulins and microtubules are subjected to several
post-translational modifications of which the reversible
detyrosination/tyrosination of the carboxy-terminal end of
most alpha-tubulins has been extensively analysed. This
modification cycle involves a specific carboxypeptidase
and the activity of the tubulin-tyrosine ligase (TTL). The
true physiological function of TTL has so far not been
established. Tubulin-tyrosine ligase (TTL) catalyses the
ATP-dependent post-translational addition of a tyrosine to
the carboxy terminal end of detyrosinated alpha-tubulin.

In normally cycling cells, the tyrosinated form of tubulin
predominates. However, in breast cancer cells, the
detyrosinated form frequently predominates, with a
correlation to tumour aggressiveness. On the other hand,
3-nitrotyrosine has been shown to be incorporated, by TTL,
into the carboxy terminal end of detyrosinated
alpha-tubulin. This reaction is not reversible by the
carboxypeptidase enzyme. Cells cultured in 3-nitrotyrosine
rich medium showed evidence of altered microtubule
structure and function, including altered cell morphology,
epithelial barrier dysfunction, and apoptosis"
/db_xref="CDD:pfam03133"

ORIGIN

Query Match 40.4%; Score 366.4; DB 10; Length 1897;
Best Local Similarity 69.3%; Pred. No. 3.4e-57;
Matches 499; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
QY 163 GTGAACCCCTCAGACCGGACATTGACCGGCTCCGGAACATCTGGATTATAAGCCCGCGCC 222
DB 572 GTGTACCCCACTAGACATGAGGGGATCAGACATCTGGATCTGGAAGCCCGAGCC 631
QY 223 AAGTCCCGGGCGGAGACATAGTGTGCATGACCGGTGTGGAGAGATCTCTGAGCTGGCA 282
DB 632 AAGTCCCGAGCGCGAGGGATTATGTGCATGAACCGCTGGATGAGATGTGAAGCTGGT 691
QY 283 GCTGCAGACACCCCTCTTCCAGGAGACAAAGTGGTGTCCAGAAGTACATCGAGACG 342
DB 692 GACTGCAACCCCATGCTCATGAAGATGGCAAGTGGATCTGCAGAAGTACATGAGCGG 751
QY 343 CCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCCTCGTCAAGGACTGG 402
DB 752 CCCCTGCTTATCTTCGGCACCAAGTTTGACCTTGAGACAGTGGTTCTCTGCTGACTGG 811
QY 403 AACCCCTGACCATCTGGTTCTACAAGGAGAGTACTTCCGTTTCTCAACTCAGCCCTTC 462
DB 812 AACCCACTCACGCTGGTGTCTACCGAGACAGTCACTTCGTTCTTCCACACAGCCCTTC 871
QY 463 TCCCTGGACAAGCTGACAGCGCCATCCACCTGTGCAACAAACCGCTCCAGAAAGTACCTG 522
DB 872 TCCCTGAAGAACCTGACAACTCTGTGCACTGTGTAACTTCATCCATCCAGAGACACTTG 931
QY 523 AAGAATGATGTGGCGCGCAGCCCCCTGTGCTCCCGCACACAAACATGTGACACACAGG 582
DB 932 GAGGCTTCCTGTGCACCGGCACCCGATGCTGCCCCAGATAACATGTGGTCCAGCCAGAG 991
QY 583 TTCCAGGAGTACTGACAGCGCCAGGGCCGTGGCGCGGTGGGGGAGCGCTCATCTACCGG 642
DB 992 TTTCCAGGCCCACTTGACGAGAGTAGATGCCCCCAAGGGCTGGTCCAGCGCTCATCGTGCCA 1051
QY 643 TCCATGAAGAGGCCATGCCCAAGGTGGCGCCAGGACCAACGCTGGAGCCCTCGC 702
DB 1052 GGCATGAAGGTGCTGTGATCCAGCCCTGACAGACCTCCCAAGACACACTGACGTGCGCCG 1111
QY 703 AAGAAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTTGTGGAGGAGACTTCAGGCCCTCG 762
DB 1112 AAGGCCAGCTTTGAGCTCTATGGGGCAGACTTTGTGTGTGGGGAAGACTTCCAGCCCTCG 1171
QY 763 CTGATGAGATCAATTCAGCCCGCCACCATGACCCCGTCCAGCCGGTCAAGCCCGAGCTG 822
DB 1172 TTGATGAATCAATGCCAGCCCGCCACCTTCCAGCGCTGTCTACGCGCCCTC 1231
QY 823 TGTGCACAGGTGCAGGAGGACACCATCAAGGTGGCGGTGGAGCCGAGCTGTGACATCGGC 882
DB 1232 TGTGCGGTGTGACGACAGATACCTCGTGTGGTTCATCGACCGGGCCCTGACCCGAGC 1291

RESULT 14

AC119959
LOCUS Mus musculus chromosome 15, clone RP24-467H19, complete sequence.
DEFINITION AC119959
ACCESSION AC119959
VERSION AC119959.8 GI:50540805
163783 bp DNA linear ROD 23-JUL-2004

KEYWORDS

SOURCE

ORGANISM

HTG.
Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 163783)

Birren, B., Nuebaum, C. and Lander, E.

Mus musculus chromosome 15, clone RP24-467H19

Unpublished

2 (bases 1 to 163783)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 163783)

Birren, B., Nusbaum, C., Lander, E., Abouellail, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., McLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (26-MAY-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 163783)

Birren, B., Nusbaum, C., Lander, E., Abouellail, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,

Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thonann, N.,
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodor, J.,
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (26-SEP-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 26, 2004 this sequence version replaced gi:52077711.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WTHR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L24098

Center clone name: 455_J_20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 82149: contig of 82149 bp in length
* 82150 82249: gap of unknown length
* 98003: contig of 15754 bp in length
* 98004 98103: gap of unknown length
* 98104 201420: contig of 103317 bp in length.

FEATURES

source

1. .201420
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-455J20"
/clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Query Match 31.1%; Score 281.6; DB 2; Length 201420;
Best Local Similarity 78.7%; Pred. No. 6.2e-42;
Matches 354; Conservative 0; Mismatches 84; Indels 12; Gaps 1;

QY 469 GACAAGCTGACAGCGCCATCCACCTGTGTGCAACACGCCGTCAGAGTACCTGAAGAAT 528
Db |||||
QY 4228 GACATTGCCACAGCGCCATCCACCTGTGTAAACAATCCATCCAGAGCGCCTCAAAAAT 4287
Db |||||
QY 529 GATGTGGCGCGACGCCCTGCTGCCCGCACACAATGTGGACGACGACGAGTTCCAG 588
Db |||||
QY 4288 GATAAAGAGCGCAGTCCGCTGCTACCTTGTATACATGTGACCACTCGTTTCCAG 4347
Db |||||
QY 589 GAGTACCTGACGCGCCAGCGCGCTGTGGGGGACGCTCATCTACCCGCTCCATG 648
Db |||||
QY 4348 GAGTACTCTGCAAGAGGGGCGCAGGAGGAACGTGGGTAGCATCATCTACCCGCTCTATG 4407
Db |||||
QY 649 AAGAAGGCCATCGCCCAACGCATGAAGTGGCGCCAGGACCACTGGAGCCTCGCAAGAAC 708
Db |||||
QY 4408 AAAAGAGCTGTCAACCAATGCCATGAGGGTAGCCAGGACCACTAGAAGCCCGTAAGAAC 4467
Db |||||
QY 709 AGCTTTGAGCTCTACGGGGCTGACTTCTCTTGGGAGGGAATTCAGGCCCTGGCTGATC 768
Db |||||
QY 4468 AGCTTCGAGCTCTACGGAGCGGACTTCTATCTGTGGGCGGAGACTTCAAGCCATGGCTTATC 4527
Db |||||
QY 769 GAGATCAATTCAGAGCCCAACCGCTCCAGCGCGGTCCAGGCCAGCTGTGTGCA 828
Db |||||
QY 4528 GAGATCAATTCAGAGCCCAACCGCTCCAGCTCCGCTCAGCTCAGCTCAGCTCTGTGCC 4587
Db |||||

QY 829 CAGGTCCAGGAGGACACCATCAAGTGGCGG-----TGGACCGCAGCTGTGAC 876
Db |||||
QY 4588 CAGGTGAGGAGGACACCATCAAGTGGTGGTGGATCGGAAACTGGACCGAAACTGTGAC 4647
Db |||||
QY 877 ATCGGCAACTTTCGAGGCTCCTGTGGAGGCGAG 906
Db |||||
QY 4648 ATCGGCAACTTTGAGCTTCTGTGGCGGCGAG 4677
Db |||||

Search completed: September 25, 2005, 11:26:56

Job time : 4207.65 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 23:48:14 ; Search time 1252.07 Seconds
(without alignments)
8666.982 Million cell updates/sec

Title: US-10-615-659-1_copy_153_1775
Perfect score: 1623
Sequence: 1 agggcatcagcatctcaaa.....cgccggagaaagatttca 1623

Scoring table: IDENTITY NUC
CAPTOP 10.0 , Capext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0 %
Maximum Match 100 %
Listing first 45 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17:	/cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18:	/cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19:	/cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20:	/cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21:	/cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22:	/cgn2_6/ptodata/2/pubpna/US10J_PUBCOMB.seq:*
23:	/cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24:	/cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1623	100.0	1838	19	US-10-615-659-1
2	1623	100.0	1838	19	US-10-615-659-1
3	1619.8	99.8	3554	19	US-10-615-659-12
4	1619.8	99.8	3554	19	US-10-615-659-12
5	1505.8	92.8	1859	19	US-10-615-659-10
6	1505.8	92.8	1859	19	US-10-615-659-10
7	1475.8	90.9	3465	19	US-10-615-659-11
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 12, Appl
					Sequence 10, Appl
					Sequence 11, Appl

8	1475.8	90.9	3465	19	US-10-635-977-11	Sequence 11, Appli
9	1474.2	90.8	1939	19	US-10-615-659-9	Sequence 9, Appli
10	1474.2	90.8	1939	19	US-10-635-977-9	Sequence 9, Appli
11	726	44.7	726	19	US-10-615-659-3	Sequence 3, Appli
12	726	44.7	726	19	US-10-635-977-3	Sequence 3, Appli
c 13	501	30.9	101270	20	US-10-723-860-631	Sequence 631, App
c 14	432.8	26.7	2380	18	US-10-275-595A-39	Sequence 39, Appli
c 15	427	26.3	190	16	US-10-029-386-24894	Sequence 24894, A
16	421.4	26.0	2358	15	US-10-102-524-1765	Sequence 1765, Ap
17	421.4	26.0	2553	21	US-10-956-157-2016	Sequence 2016, Ap
18	421.4	26.0	2553	22	US-10-756-149-1371	Sequence 1971, Ap
19	421.4	26.0	3001	15	US-10-037-270-494	Sequence 494, App
20	421.4	26.0	3001	17	US-10-117-722-494	Sequence 494, App
21	418	25.8	2848	17	US-10-108-260A-1766	Sequence 1766, Ap
22	413	25.4	2611	9	US-09-925-298-199	Sequence 199, App
23	413	25.4	2611	14	US-10-102-806-199	Sequence 199, App
c 24	387.6	23.9	526	13	US-10-027-632-135927	Sequence 135927,
c 25	387.6	23.9	526	13	US-10-027-632-135928	Sequence 135928,
c 26	387.6	23.9	526	17	US-10-027-632-135927	Sequence 135927,
c 27	387.6	23.9	526	17	US-10-027-632-135928	Sequence 135928,
28	381.8	23.5	1400	21	US-10-956-157-9356	Sequence 9356, Ap
29	381.8	23.5	1673	21	US-10-956-157-4121	Sequence 4121, Ap
30	319.8	19.7	1400	21	US-10-956-157-7251	Sequence 7251, Ap
31	196	12.1	755	13	US-10-027-632-135929	Sequence 135929,
32	196	12.1	755	13	US-10-027-632-135930	Sequence 135930,
33	196	12.1	755	17	US-10-027-632-135929	Sequence 135929,
34	196	12.1	755	17	US-10-027-632-135930	Sequence 135930,
35	190.6	11.7	1958	21	US-10-956-157-4903	Sequence 4903, Ap
36	166	10.2	1728	22	US-10-450-763-5362	Sequence 5362, Ap
37	164.4	10.1	2250	22	US-10-450-763-5362	Sequence 5362, Ap
38	163.4	10.1	418	9	US-09-983-965-2032	Sequence 2032, Ap
39	131.8	8.1	492	10	US-09-918-995-3253	Sequence 3253, A
c 40	130	8.0	525	16	US-10-029-386-11174	Sequence 11174, A
41	122.6	7.6	436	17	US-10-242-535A-7660	Sequence 7660, Ap
42	122.6	7.6	436	18	US-10-085-783A-7660	Sequence 7660, Ap
43	120.4	7.4	2241	19	US-10-615-659-27	Sequence 27, Appl
44	120.4	7.4	2241	19	US-10-635-977-27	Sequence 27, Appl
45	113.4	7.0	279	18	US-10-424-599-119452	Sequence 119452,

ALIGNMENTS

RESULT 1
US-10-615-659-1
; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; TYPE: DNA
; LENGTH: 1838
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1

Query Match 100.0%; Score 1623; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGGCATCAGCATCTCAAGTGGTGGTCAGCCACAGAGCTGCAGCAGCAGCAGA 60
|||||

Db 153 ATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACGAGCTGCAGGAGCAGCAGA 212
QY 61 AGCAAGCCAGGACACAGAGGAGAGGCGCGGAGCAGCGACCTCTGAGCAGCAGGCAAGAT 120
Db 213 AGCAAGCCAGGACACAGAGGAGAGGCGCGGAGCAGCGACCTCTGAGCAGCAGGCAAGAT 272
QY 121 GCTGAAATGCTGAGGCAAGCTCAGGGGCTCTCCCGGGCAGCTTGTGACATCGCGTGC 180
Db 273 GCTGAAATGCTGAGGCAAGCTCAGGGGCTCTCCCGGGCAGCTTGTGACATCGCGTGC 332
QY 181 AAGGTGTGCAGGCTTACCTGGGGCAGCTGAGCATGAGGACATCGACACGTCAGCAGAT 240
Db 333 AAGGTGTGCAGGCTTACCTGGGGCAGCTGAGCATGAGGACATCGACACGTCAGCAGAT 392
QY 241 GCCGTGGAGGACCTCACTCAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 300
Db 393 GCCGTGGAGGACCTCACTCAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 452
QY 301 GTTCATGCGGATGCTTTCACTCTCCAATTCAAGAAATTACTTTTCGAGTGCCAGGCTGTG 360
Db 453 GTTCATGCGGATGCTTTCACTCTCCAATTCAAGAAATTACTTTTCGAGTGCCAGGCTGTG 512
QY 361 CTGCAATAGAAATCAGCTGTGAACTCTCAGACGACATTTGACGGGCTCCGGACATCTGG 420
Db 513 CTGCAATAGAAATCAGCTGTGAACTCTCAGACGACATTTGACGGGCTCCGGACATCTGG 572
QY 421 ATTATTAAGCCCGGCGCAAGTCCCGGGCCGAGACATAGTGTGATGACACCGTGTGGAG 480
Db 573 ATTATTAAGCCCGGCGCAAGTCCCGGGCCGAGACATAGTGTGATGACACCGTGTGGAG 632
QY 481 GAGATCTCGAGTGTGACGTGCAGACCACTCTTTCCAGGGACAAAGTGGGTGTGTC 540
Db 633 GAGATCTCGAGTGTGACGTGCAGACCACTCTTTCCAGGGACAAAGTGGGTGTGTC 692
QY 541 CAGAAGTACATCGAGACGCGCTGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
Db 693 CAGAAGTACATCGAGACGCGCTGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 752
QY 601 TTCTCTCGTCAGGACTGGAAACCCCTGACCATCTGGTCTTCAAGGAGAGTTACTTTCGG 660
Db 753 TTCTCTCGTCAGGACTGGAAACCCCTGACCATCTGGTCTTCAAGGAGAGTTACTTTCGG 812
QY 661 TTCTCAACTCAGGCTTTCTCTTGACAAGCTGGACAGCGCCATCCACTGTGCAACAC 720
Db 813 TTCTCAACTCAGGCTTTCTCTTGACAAGCTGGACAGCGCCATCCACTGTGCAACAC 872
QY 721 GCGCTCCAGAAGTACTGGAAGATGATGTGGCGCAGCGCCCTGCTGCGCCGACACAC 780
Db 873 GCGCTCCAGAAGTACTGGAAGATGATGTGGCGCAGCGCCCTGCTGCGCCGACACAC 932
QY 781 ATGTGGACACGACACAGGTTCCAGAGTACCTGCAGCGCCAGGCGCTGGCGCGTGTGG 840
Db 933 ATGTGGACACGACACAGGTTCCAGAGTACCTGCAGCGCCAGGCGCTGGCGCGTGTGG 992
QY 841 GCGAGGTCATCTACCCGTCCATGAAGAGCCATCGCCACGCCATGAAGTGGCCCCAG 900
Db 993 GCGAGGTCATCTACCCGTCCATGAAGAGCCATCGCCACGCCATGAAGTGGCCCCAG 1052
QY 901 GACCGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTCTGG 960
Db 1053 GACCGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTCTGG 1112
QY 961 AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCCGTCCAG 1020
Db 1113 AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCCGTCCAG 1172
QY 1021 CCGGTACCGGCCAGCTGTGTGCAAGGTGCAAGGAGCACCATCAAGGTGGCCGTGGAC 1080
Db 1173 CCGGTACCGGCCAGCTGTGTGCAAGGTGCAAGGAGCACCATCAAGGTGGCCGTGGAC 1232
QY 1081 CGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGGAGCAGCCGCTGGTTGAGCGGCC 1140
Db 1233 CGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGGAGCAGCCGCTGGTTGAGCGGCC 1292

QY 1141 CCATTTCAGCGGTCCGACCTCTCGTGGCGGCGTCACTGTGAGGAGCCAGGAGGCAG 1200
Db 1293 CCATTTCAGCGGTCCGACCTCTCGTGGCGGCGTCACTGTGAGGAGCCAGGAGGCAG 1352
QY 1201 GTGTGCCCCGTCTGCAACCTCAAGGCTCGGCTCGCTGTGGAACGCGCAGCGCTGAAG 1260
Db 1353 GTGTGCCCCGTCTGCAACCTCAAGGCTCGGCTCGCTGTGGAACGCGCAGCGCTGAAG 1412
QY 1261 GCACGGGGCCCTCGGCGCATGCAGACCTGCCAGGACCCCATCACAGCTCTCCAG 1320
Db 1413 GCACGGGGCCCTCGGCGCATGCAGACCTGCCAGGACCCCATCACAGCTCTCCAG 1472
QY 1321 CGGGACTTTGGGACTGAAGGAGAGAGGGGCTCCCCCTGTGGCCTTGTGTGCACCCCTTAAAG 1380
Db 1473 CGGGACTTTGGGACTGAAGGAGAGAGGGGCTCCCCCTGTGGCCTTGTGTGCACCCCTTAAAG 1532
QY 1381 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGTCTGGGAAGGTG 1440
Db 1533 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGTCTGGGAAGGTG 1592
QY 1441 GAGCTCCCGGCTGCCCTGTCCGACAGTGGACAGTCAAGGCCCCAAACACCGGTGTCCCC 1500
Db 1593 GAGCTCCCGGCTGCCCTGTCCGACAGTGGACAGTCAAGGCCCCAAACACCGGTGTCCCC 1652
QY 1501 GTAGCCCAAGCCCGCAAAAGCTGGGATCCAAACCAAGTAAATGCGCACCCCGCTGGAGCCT 1560
Db 1653 GTAGCCCAAGCCCGCAAAAGCTGGGATCCAAACCAAGTAAATGCGCACCCCGCTGGAGCCT 1712
QY 1561 GTGTGCGGGGCTGAAGACAGACAGAGGCGCGCTGCGTCCGCGCCGGAGGAAAAGGT 1620
Db 1713 GTGTGCGGGGCTGAAGACAGACAGAGGCGCGCTGCGTCCGCGCCGGAGGAAAAGGT 1772
QY 1621 TCA 1623
Db 1773 TCA 1775

RESULT 2

US-10-635-977-1
; Sequence 1, Application US/10635977
; Publication NO. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-635-977-1

Query Match 100.0%; Score 1623; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACGAGCTGCAGGAGCAGCAGA 60
Db 153 ATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACGAGCTGCAGGAGCAGCAGA 212
QY 61 AGCAAGCCAGGACACAGGAGGAGGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGAT 120

Db	1293	CCATTACAGCGGTCCGACCTCTCGTGTGGCGGGCGTCAGTGTGAGAGAGACCGACGAGGCAG	1352
Qy	1201	GTGCTGCCCGTCTGCAACCTCAAGGGCTCGGCCCTCGCTGTTGGACGCGCAGCCGCTGAAG	1260
Db	1353	GTGCTGCCCGTCTGCAACCTCAAGGGCTCGGCCCTCGCTGTTGGACGCGCAGCCGCTGAAG	1412
Qy	1261	GCACGGGGCCCCCTGGCCATGCCAGACCCCTGCCACGGGACCCCCATCACACAGCTCTCCAG	1320
Db	1413	GCACGGGGCCCCCTGGCCATGCCAGACCCCTGCCACGGGACCCCCATCACACAGCTCTCCAG	1472
Qy	1321	CGGACCTTGGGACTGAAGGAAGAGAGGGGCTCCCCCTGGCCCTTGCTGGCACCCCTTAAGG	1380
Db	1473	CGGACCTTGGGACTGAAGGAAGAGAGGGGCTCCCCCTGGCCCTTGCTGGCACCCCTTAAGG	1532
Qy	1381	GGGGCAGCCAGAGCGGTGGAGCGGCACAGCCACCGGCACCAAGCTGCTGGGAGGGTG	1440
Db	1533	GGGGCAGCCAGAGCGGTGGAGCGGCACAGCCACCGGCACCAAGCTGCTGGGAGGGTG	1592
Qy	1441	GAGCTCCGGCCCTGCCCTCTGCGCCACGTGGACAGTCAGCGCCCAACACCGGTGTCGCC	1500
Db	1593	GAGCTCCGGCCCTGCCCTCTGCGCCACGTGGACAGTCAGCGCCCAACACCGGTGTCGCC	1652
Qy	1501	GTAGCCCAGCCCGCAAAAGCTGGATCCAAACAGCTAAATGGCACCCCGCTGGAGCCT	1560
Db	1653	GTAGCCCAGCCCGCAAAAGCTGGATCCAAACAGCTAAATGGCACCCCGCTGGAGCCT	1712
Qy	1561	GTGCTGGGGGGCTGAAGACAGACAGAGGGGCGCGCTGCGTCGGCGCCCGGAGGAAAGGT	1620
Db	1713	GTGCTGGGGGGCTGAAGACAGACAGAGGGGCGCGCTGCGTCGGCGCCCGGAGGAAAGGT	1772
Qy	1621	TCA 1623	
Db	1773	TCA 1775	

RESULT 3

Query Match	99.8%	Score 1619.8	DB 19	Length 3554
Best Local Similarity	99.9%	Pred. No. 0		
Matches 1621	Conservative 0	Mismatches 2	Indels 0	Gaps 0
Qy 1	ATGGCATCCAGCATCTCTCAAGTGGTGGTTCAGGCCACAGAGCTGCAGCAGCGAGCAGCAGA	60		
Db 1695	ATGGCATCCAGCATCTCTCAAGTGGTGGTTCAGGCCACAGAGCTGCAGCAGCGAGCAGCAGA	1754		
Qy 61	AGCAAGCCCAAGGACCAAGAGGGAGGAGCGCGGAGCAGCGCACTGTAGCAGCAGCGCAAGAT	120		
Db 1755	AGCAAGCCCAAGGACCAAGAGGGAGGAGCGCGGAGCAGCGCACTGTAGCAGCAGCGCAAGAT	1814		
Qy 121	GCTGAAATCTCAGCGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC	180		
Db 1815	GCTGAAATCTCAGCGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC	1874		
Qy 181	AAGGTGTGCCAGGCGCTACCTCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT	240		

[illegible]

Db	2955	GCACGGGGCCCCCTCGGCCATCGCAGACCCCTGCGCACGGGACCCCCATATCACAGCTCTCCAG	3014
Qy	1321	CGGACCTTGGGACTGAAGGAAGAGAGGGGCTCCCCCTGGCCTTGTGTGGCACCCCTTAAGG	1380
Db	3015	CGGACTTGGGACTGAAGGAAGAGAGGGGCTCCCCCTGGCCTTGTGTGGCACCCCTTAAGG	3074
Qy	1381	GGGGCAGCCGAGAGCGGTGGAGCCGCACACGCCACCCGCGACCAAGAGCTGTGGGAAGGTG	1440
Db	3075	GGGGCAGCCGAGAGCGGTGGAGCCGCACACGCCACCCGCGACCAAGAGCTGTGGGAAGGTG	3134
Qy	1441	GAGTCCCGGCTGTCCTCTGCGCACGTTGGACAGTCAGGCCCCCAACACCGGTGTCCCC	1500
Db	3135	GAGTCCCGGCTGTCCTCTGCGCACGTTGGACAGTCAGGCCCCCAACACCGGTGTCCCC	3194
Qy	1501	GTAGCCAGCCCGCAAAAGCTGGATCCAAAACAGCTAAATCGCACCCCGCTGGAGCCT	1560
Db	3195	GTAGCCAGCCCGCAAAAGCTGGATCCAAAACAGCTAAATCGCACCCCGCTGGAGCCT	3254
Qy	1561	GTGCTCGGGGCTGAAGACAGCAGAGGGCGCTCGCTCCGCGCCGCGAGGAAAAGGT	1620
Db	3255	GTGCTCGGAGCCTGAAGACAGCAGAGGGCGCTCGCTCCGCGCCGCGAGGAAAAGGT	3314
Qy	1621	TCA 1623	
Db	3315	TCA 3317	
RESULT 4			
US-10-635-977-12			
; Sequence 12, Application US/10635977			
; Publication No. US20040171131A1			
; GENERAL INFORMATION:			
; APPLICANT: Bristol-Myers Squibb Company			
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN			
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42			
; FILE REFERENCE: D028JA CIP			
; CURRENT APPLICATION NUMBER: US/10/635,977			
; CURRENT FILING DATE: 2003-08-07			
; PRIOR APPLICATION NUMBER: U.S. 60/394,725			
; PRIOR FILING DATE: 2002-07-09			
; PRIOR APPLICATION NUMBER: U.S.10/615,659			
; PRIOR FILING DATE: 2003-07-09			
; NUMBER OF SEQ ID NOS: 103			
; SOFTWARE: PatentIn version 3.2			
; SEQ ID NO 12			
; LENGTH: 3554			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-635-977-12			
Query Match 99.8%; Score 1619.8; DB 19; Length 3554;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 1621; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	1	ATGCGATCCAGCATCCTCAAGTCGGTGGTCAGCCACAGAGCTGCAGCAGGAGCAGCAGA	60
Db	1695	ATGCGATCCAGCATCCTCAAGTCGGTGGTCAGCCACAGAGCTGCAGCAGGAGCAGCAGA	1754
Qy	61	AGCAAGCCAGGACCCAGGAGGAGGCGCGGAGAGCGACCTGTAGCAGCAGCAGCAT	120
Db	1755	AGCAAGCCAGGACCCAGGAGGAGGCGCGGAGAGCGACCTGTAGCAGCAGGCAAGAT	1814
Qy	121	GCTGAAATGCTCAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGCTGTGGACATCGCGTGC	180
Db	1815	GCTGAAATGCTCAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGCTGTGGACATCGCGTGC	1874
Qy	181	AAGGTGTGCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT	240
Db	1875	AAGGTGTGCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT	1934
Qy	241	GCGTGGAGACCTTACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC	300
Db	1935	GCGTGGAGACCTTACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC	1994

QY 301 GTTCATGGCGATGTTTCATCTCCAAATTCAGAAATTAATTTTCGAGTGCAGGCTCTG 360
DB 1995 GTTCATGGCGATGTTTCATCTCCAAATTCAGAAATTAATTTTCGAGTGCAGGCTCTG 2054
QY 361 CTGAATAGAAATACAGTCTGTGAACCCCTCAGACGGACATTCACGGGCTCCGGAACATCTGG 420
DB 2055 CTGAATAGAAATACAGTCTGTGAACCCCTCAGACGGACATTCACGGGCTCCGGAACATCTGG 2114
QY 421 ATTATAAAGCCCGCGGCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 480
DB 2115 ATTATAAAGCCCGCGGCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 2174
QY 481 GAGATCTGAGCTGGCAGCTGCAGACCCCTCTTTCAGAGGACACAAAGTGGGTGCTC 540
DB 2175 GAGATCTGAGCTGGCAGCTGCAGACCCCTCTTTCAGAGGACACAAAGTGGGTGCTC 2234
QY 541 CAGAAGTACATCAGACGCCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
DB 2235 CAGAAGTACATCAGACGCCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 2294
QY 601 TTCTCTGTCAGGATGGAACCCCTGACCATCTGTTTCAAGAGAGATTACTTTCGG 660
DB 2295 TTCTCTGTCAGGATGGAACCCCTGACCATCTGTTTCAAGAGAGATTACTTTCGG 2354
QY 661 TTCTCAACTCAGGCTTCTCCTTGGACAGCTGGAAGCCCATCACCCTGTGCAACAAC 720
DB 2355 TTCTCAACTCAGGCTTCTCCTTGGACAGCTGGAAGCCCATCACCCTGTGCAACAAC 2414
QY 721 GCGGTCCAGAAATGATGTGGCGGACGCCCTGCTGCCCGCACACAAC 780
DB 2415 GCGGTCCAGAAATGATGTGGCGGACGCCCTGCTGCCCGCACACAAC 2474
QY 781 ATGTGACACAGCACAGGTTCCAGGAGTACCTGTCAGCGCCAGGCGCGTGGCGCGTGTGG 840
DB 2475 ATGTGACACAGCACAGGTTCCAGGAGTACCTGTCAGCGCCAGGCGCGTGGCGCGTGTGG 2534
QY 841 GGCAGGTGATCTACCGTCCATGAAGAGGCCATATGCCCCACGCCATGAAGTGGCCAG 900
DB 2535 GGCAGGTGATCTACCGTCCATGAAGAGGCCATATGCCCCACGCCATGAAGTGGCCAG 2594
QY 901 GACCACTGGAGCCTCGAAGAACAGCTTTGAGTCTTACGGGCTGACTTCGTCCTTGG 960
DB 2595 GACCACTGGAGCCTCGAAGAACAGCTTTGAGTCTTACGGGCTGACTTCGTCCTTGG 2654
QY 961 AGGGAATTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCACCATGCAACCGTCCAG 1020
DB 2655 AGGGAATTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCACCATGCAACCGTCCAG 2714
QY 1021 CCGGTCAACCGCCAGCTGTGTGACAGGTGCAAGAGGACACCATCAAGTGGCCGTGGAC 1080
DB 2715 CCGGTCAACCGCCAGCTGTGTGACAGGTGCAAGAGGACACCATCAAGTGGCCGTGGAC 2774
QY 1081 CGCAGCTGTGACATCGCAACTTCGAGCTCTGTGGAGCGCGGCTGGTGGAGCGCCCC 1140
DB 2775 CGCAGCTGTGACATCGCAACTTCGAGCTCTGTGGAGCGCGGCTGGTGGAGCGCCCC 2834
QY 1141 CCATTACGCGGCTCCGACCTCTGCTGGCGGGCTCAGTGTGAGAGAGCCAGAGGCGAG 1200
DB 2835 CCATTACGCGGCTCCGACCTCTGCTGGCGGGCTCAGTGTGAGAGAGCCAGAGGCGAG 2894
QY 1201 GTGCTGCCCTGTGCAACCTCAAGGCTTCGGCTTCGCTGTGTGACGCGACGCGCTGAAG 1260
DB 2895 GTGCTGCCCTGTGCAACCTCAAGGCTTCGGCTTCGCTGTGTGACGCGACGCGCTGAAG 2954
QY 1261 GCACGGGGCCCTCGGCCATGCGCAGACCTTCCAGGAGCCCCCATCAGCAGCTCCAG 1320
DB 2955 GCACGGGGCCCTCGGCCATGCGCAGACCTTCCAGGAGCCCCCATCAGCAGCTCCAG 3014
QY 1321 CGGCACTTGGGAATGAAGGAAGAGAGGGGCTCCCTCTGCGCTTGTGGCACCTTTAAGG 1380
DB 3015 CGGCACTTGGGAATGAAGGAAGAGAGGGGCTCCCTCTGCGCTTGTGGCACCTTTAAGG 3074

QY 1381 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGCTGGGAAGGTG 1440
DB 3075 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAAAGCTGCTGGGAAGGTG 3134
QY 1441 GAGCTCCCGGCTGCGCCCTGTCGCCACAGTGGAGCAGTCAGGCCCAAAACACCGGTGTCCCC 1500
DB 3135 GAGCTCCCGGCTGCGCCCTGTCGCCACAGTGGAGCAGTCAGGCCCAAAACACCGGTGTCCCC 3194
QY 1501 GTAGCCACGCGCCCAAAAGCTGGGATCCAAACCAAGCTAAATCGGACCCGCTGGAGCCT 1560
DB 3195 GTAGCCACGCGCCCAAAAGCTGGGATCCAAACCAAGCTAAATGAGCACCCGCTGGAGCCT 3254
QY 1561 GTGCTCGGCGCTGAAGACAGCAGAGGGCGCTGCTCCGCGCCCGGAGGAAAGGT 1620
DB 3255 GTGCTCGGAGCCTGAAGACAGCAGAGGGCGCTGCTCCGCGCCCGGAGGAAAGGT 3314
QY 1621 TCA 1623
DB 3315 TCA 3317

RESULT 5
US-10-615-659-10
; Sequence 10, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-10

Query Match 92.8%; Score 1505.8; DB 19; Length 1859;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1574; Conservative 0; Mismatches 2; Indels 47; Gaps 2;

QY 1 ATGGCATCCAGCATCTCAAGTGGGTGGTTCAGCCACAGAGCTGCAGAGGAGCAGCAGA 60
DB 47 ATGGCATCCAGCATCTCAAGTGGGTGGTTCAGCCACAGAGCTGCAGAGGAGCAGCAGA 106
QY 61 AGCAAGCCACAGGACCCAGAGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGAT 120
DB 107 AGCAAGCCACAGGACCCAGAGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGAT 166
QY 121 GCTGAAAATGCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGCGTGC 180
DB 167 GCTGAAAATGCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGCGTGC 226
QY 181 AAGGTGTGCGAGCCCTTACCTGGGCGAGCTGGAGCATGAGGACATCCACACGTCAGCAGAT 240
DB 227 AAGGTGTGCGAGCCCTTACCTGGGCGAGCTGGAGCATGAGGACATCCACACGTCAGCAGAT 286
QY 241 GCCGTGGAGGACCTCACTCAGGCGCAGTGGAGGAGCTTGCACCCAGCAGTACTACTCCCTC 300
DB 287 GCCGTGGAGGACCTCACTCAGGCGCAGTGGAGGAGCTTGCACCCAGCAGTACTACTCCCTC 346
QY 301 GTTCATGGCGATGCTTTTCAATTCAGAAATTAATTTTCGAGTGCAGGCTCTG 360
DB 347 GTTCA-----TGCAGGCTCTG 363
QY 361 CTGAATAGAAATCAGCTCTGTGAACCCCTCAGAGCGGACATTTGACGGGCTCCGGGAACATCTGG 420
DB 364 CTGAATAGAAATCAGCTCTGTGAACCCCTCAGAGCGGACATTTGACGGGCTCCGGGAACATCTGG 423


```
QY 421 ATTATAAAGCCCGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGGAACGCTGTGGAG 480
|
|
|
Db 424 ATTATAAAGCCCGCCCAAGTCCCGGGCCGAGACA TAGTGTGCATGGACCGTGTGGAG 483
|
|
|
QY 481 GAGATCCTGGAGTGGCAGTGCAGACCAACCTCTTTCCAGGGACAACAAGTGGGTGTC 540
|
|
|
Db 484 GAGATCCTGGAGTGGCAGTGCAGACCAACCTCTTTCCAGGGACAACAAGTGGGTGTC 543
|
|
|
QY 541 CAGAAGTACATCGAGACGGCGTGTCTATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
|
|
|
Db 544 CAGAAGTACATCGAGACGGCGTGTCTATCTGTGACACCAAGTTCGACATCAGACAGTGG 603
|
|
|
QY 601 TTCTCTCGTCA CGGAC TGGAA CCCCCTTGACCATCTGTTCTTACAAGGAGATTACTTGGG 660
|
|
|
Db 604 TTCTCTCGTCA CGGAC TGGAA CCCCCTTGACCATCTGTTCTTACAAGGAGATTACTTGGG 663
|
|
|
QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 720
|
|
|
Db 664 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 723
|
|
|
QY 721 GCCGTCCAGAAGTACCTGAAGAATGATGTGGCGCGCAGCCCTGCTGCGCCGACACAAC 780
|
|
|
Db 724 GCCGTCCAGAAGTACCTGAAGAATGATGTGGCGCGCAGCCCTGCTGCGCCGACACAAC 783
|
|
|
QY 781 ATGTGGACCAAGCACCAGGTTTCCAGGAGTACCTGACGCGCCAGGGCGTGGCGCGTGTGG 840
|
|
|
Db 784 ATGTGGACCAAGCACCAGGTTTCCAGGAGTACCTGACGCGCCAGGGCGTGGCGCGTGTGG 843
|
|
|
QY 841 GGCAGGCTCATCTACCCGTCTCATGAAGAAGGCATCGCCCCACGCCATGAAGTGGCCGAC 900
|
|
|
Db 844 GGCAGGCTCATCTACCCGTCTCATGAAGAAGGCATCGCCCCACGCCATGAAGTGGCCGAC 903
|
|
|
QY 901 GACCACTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCTGTGG 960
|
|
|
Db 904 GACCACTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCTGTGG 963
|
|
|
QY 961 AGGGACTTCAGGCGCTTGGGTGATCGAGATCAATTCAGCGCCCAACCATGACCCGTCACG 1020
|
|
|
Db 964 AGGGACTTCAGGCGCTTGGGTGATCGAGATCAATTCAGCGCCCAACCATGACCCGTCACG 1023
|
|
|
QY 1021 CGGCTCACGGCCAGCTGTGTGACAGGTGCAGAGGACACCATCAAGTGGCGGTGGAC 1080
|
|
|
Db 1024 CGGCTCACGGCCAGCTGTGTGACAGGTGCAGAGGACACCATCAAGTGGCGGTGGAC 1083
|
|
|
QY 1081 CGCAGCTGTGACATCGCAACTTCGAGCTCTGTGTGAGGACGCGGTGTGTGAGCGGCC 1140
|
|
|
Db 1084 CGCAGCTGTGACATCGCAACTTCGAGCTCTGTGTGAGGACGCGGTGTGTGAGCGGCC 1139
|
|
|
QY 1141 CCATTACGCGGGTCCGACCTCTCGTGGCGGGCGTCAAGTGTGAGGAGAGCCAGGAGGCAG 1200
|
|
|
Db 1140 CCATTACGCGGGTCCGACCTCTCGTGGCGGGCGTCAAGTGTGAGGAGAGCCAGGAGGCAG 1199
|
|
|
QY 1201 GTGTGCTCCGCTGTGCAACTCAAGGCTCGGCTCTGCTTTGGAACGCGAGCGCTGAAG 1260
|
|
|
Db 1200 GTGTGCTCCGCTGTGCAACTCAAGGCTCGGCTCTGCTTTGGAACGCGAGCGCTGAAG 1259
|
|
|
QY 1261 GCACGGGGCCCTCGGCCATGCAGACCTGCGCCAGGGACCCCAACACAGCTCTCCAG 1320
|
|
|
Db 1260 GCACGGGGCCCTCGGCCATGCAGACCTGCGCCAGGGACCCCAACACAGCTCTCTCAG 1319
|
|
|
QY 1321 CGGGACTTGGGACTGAAGGAAGAGGGGCTCCCTCTGGCCCTTGTGTGACCCCTTAAGG 1380
|
|
|
Db 1320 CGGGACTTGGGACTGAAGGAAGAGGGGCTCCCTCTGGCCCTTGTGTGACCCCTTAAGG 1379
|
|
|
QY 1381 GGGGCHAGCCGAGAGCGGTGGAGCCGACACGCCACCCGACCAAAAGCTCTGGGAAGGTG 1440
|
|
|
Db 1380 GGGGCHAGCCGAGAGCGGTGGAGCCGACACGCCACCCGACCAAAAGCTCTGGGAAGGTG 1439
|
|
|
QY 1441 GAGCTCCCGGCTGCGCCCTGTGCGCACGTGGACAGTCAAGGCGCCCAAAACACCGGTGTC 1500
|
|
|
Db 1440 GAGCTCCCGGCTGCGCCCTGTGCGCACGTGGACAGTCAAGGCGCCCAAAACACCGGTGTC 1499
|
|
|
```

```
QY 1501 GTAGCCCAAGCCCGCCAAAGCTGGGATCCAAACCAAGCTAAATGCGCACCCCGCTGGAGCCT 1560
|
|
|
Db 1500 GTAGCCCAAGCCCGCCAAAGCTGGGATCCAAACCAAGCTAAATGAGCACCCGCTGGAGCCT 1559
|
|
|
QY 1561 GTGCTCGGGGCTGAGACAGCAGAGGGCGGCTGCGTCCGCGCCCGGAGGAAAGCT 1620
|
|
|
Db 1560 GTGCTCGGGGCTGAGACAGCAGAGGGCGGCTGCGTCCGCGCCCGGAGGAAAGGT 1619
|
|
|
QY 1621 TCA 1623
|
|
|
Db 1620 TCA 1622
|
|
|
RESULT 6
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-10
Query Match 92.8%; Score 1505.8; DB 19; Length 1859;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1574; Conservative 0; Mismatches 2; Indels 47; Gaps 2;
QY 1 ATGGCATCCAGCATCCTCAAGTGGGTGCTCAGCCACAGAGCTGCAGAGGAGCAGCAG 60
|
|
|
Db 47 ATGGCATCCAGCATCCTCAAGTGGGTGCTCAGCCACAGAGCTGCAGAGGAGCAGCAG 106
|
|
|
QY 61 AGCAAGCCCAAGGACCCAGAGGAGGAGGCGGGAGCAGCGACTTGAGCAGAGGCAAGAT 120
|
|
|
Db 107 AGCAAGCCCAAGGACCCAGAGGAGGAGGCGGGAGCAGCGACTTGAGCAGAGGCAAGAT 166
|
|
|
QY 121 GCTGAAAATGCTGAGGCAAAAGCTCAGGGGCTTCCCGGGGAGCTTGTGGACATCGCGTGC 180
|
|
|
Db 167 GCTGAAAATGCTGAGGCAAAAGCTCAGGGGCTTCCCGGGGAGCTTGTGGACATCGCGTGC 226
|
|
|
QY 181 AAGGTGTCGACGGCTTACTGCGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 240
|
|
|
Db 227 AAGGTGTCGACGGCTTACTGCGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 286
|
|
|
QY 241 GCGGTGGAGGACCTCACTGAGGCCAGTGGGAGGACCTTGACCCAGCAGTACTACTCCCTC 300
|
|
|
Db 287 GCGGTGGAGGACCTCACTGAGGCCAGTGGGAGGACCTTGACCCAGCAGTACTACTCCCTC 346
|
|
|
QY 301 GTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAAATTACTTTTTCGAGTGCAGGCTCTG 360
|
|
|
Db 347 GTTCA-----TGCCAGGCTCTG 363
|
|
|
QY 361 CTGAATAGATCACTGTGTGAACCCCTCAGACGGACATGACGGCTCCGGACATCTGG 420
|
|
|
Db 364 CTGAATAGATCACTGTGTGAACCCCTCAGACGGACATGACGGCTCCGGACATCTGG 423
|
|
|
QY 421 ATTATAAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGATGGACCGTGTGGAG 480
|
|
|
Db 424 ATTATAAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGATGGACCGTGTGGAG 483
|
|
|
QY 481 GAGATCCCTGGAGCTGGCAGCTGCAACCAACCTCTTTTCCAGGGACAACAAGTGGGTGTC 540
|
|
|
```


Db 484 GAGATCTTGAGCTGGGAGCTGCAGACCACTCTTTCCAGGGGACAAAGTGGGTGGTC 543
Qy |||||
Db 541 CAGAAGTACATCGAGACGCCGCTCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG 600
Qy |||||
Db 544 CAGAAGTACATCGAGACGCCGCTCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG 603
Qy |||||
Db 601 TTCTCTGTCAGGACTGGAAACCCCTGACCATCTGGTTCTACAAGGAGATTACTTGGCG 660
Qy |||||
Db 604 TTCTCTGTCAGGACTGGAAACCCCTGACCATCTGGTTCTACAAGGAGATTACTTGGCG 663
Qy |||||
Db 661 TTCTCAACTCAGGCTTCTCCCTGGACAAAGCTGACAGCGCCATCCACCTGTGCAACAAC 720
Qy |||||
Db 664 TTCTCAACTCAGGCTTCTCCCTGGACAAAGCTGACAGCGCCATCCACCTGTGCAACAAC 723
Qy |||||
Db 721 GCGCTCAGAACTACCTGAAGAATGATGTGGGCGCAGCCCTCTGTCGCCGACACAAC 780
Qy |||||
Db 724 GCGCTCAGAACTACCTGAAGAATGATGTGGGCGCAGCCCTCTGTCGCCGACACAAC 783
Qy |||||
Db 781 ATGTGGACAGCAGGCTTCCAGGAGTACTGTGACGCCAGGCCGCTGGCGCGTGTGG 840
Qy |||||
Db 784 ATGTGGACAGCAGGCTTCCAGGAGTACTGTGACGCCAGGCCGCTGGCGCGTGTGG 843
Qy |||||
Db 841 GGCAGCGTCACTACCGCTCATGAAGAGCCCATCGCCACGCGCATGAAGGTGGCCAG 900
Qy |||||
Db 844 GGCAGCGTCACTACCGCTCATGAAGAGCCCATCGCCACGCGCATGAAGGTGGCCAG 903
Qy |||||
Db 901 GACCAGTGTGAGGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTTCGCTTGGG 960
Qy |||||
Db 904 GACCAGTGTGAGGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTTCGCTTGGG 963
Qy |||||
Db 961 AGGAGCTTACAGGCGCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCGTCACG 1020
Qy |||||
Db 964 AGGAGCTTACAGGCGCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCGTCACG 1023
Qy |||||
Db 1021 CCGGTACAGCCGCTGTGTGACAGCTGTCAGAGTCCAGAGGACACCATCAAGTGGCCGTGGAC 1080
Qy |||||
Db 1024 CCGGTACAGCCGCTGTGTGACAGTGTGACAGTGTGACAGGACACCATCAAGTGGCCGTGGAC 1083
Qy |||||
Db 1081 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCGCGGTGGTTGAGCCGCCC 1140
Qy |||||
Db 1084 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGG---AGCCGGTGGTTGAGCCGCCC 1139
Qy |||||
Db 1141 CCAATTACAGCGGCTCGACCTCTCGTGGCGGGGCTCAGTGTGAGGAGAGCCAGAGGAGCAG 1200
Qy |||||
Db 1140 CCAATTACAGCGGCTCGACCTCTCGTGGCGGGGCTCAGTGTGAGGAGAGCCAGAGGAGCAG 1199
Qy |||||
Db 1201 GTGCTGCCGCTGTGCAACCTCAAGGCTCGGCTCGCTGTGTCAGCGCAGCGCTGAAG 1260
Qy |||||
Db 1200 GTGCTGCCGCTGTGCAACCTCAAGGCTCGGCTCGCTGTGTCAGCGCAGCGCTGAAG 1259
Qy |||||
Db 1261 GCACGGGGCCCTCGGCCATGCCAGACCTTCGCCAGGAGCCGCCCATCACCAGCTCTCCAG 1320
Qy |||||
Db 1260 GCACGGGGCCCTCGGCCATGCCAGACCTTCGCCAGGAGCCGCCCATCACCAGCTCTCCAG 1319
Qy |||||
Db 1321 CGGGAATTGGGACTGAAGGAAGAGAGGGGCTCCCTCGCTGGCTTGTGTGGCACTTAAGG 1380
Qy |||||
Db 1320 CGGGAATTGGGACTGAAGGAAGAGAGGGGCTCCCTCGCTGGCTTGTGTGGCACTTAAGG 1379
Qy |||||
Db 1381 GGGGCAAGCCAGAGCGGTGAGCGGACAGCCAGCCAGCCAGCAAAAGCTGTGGGAAGGTG 1440
Qy |||||
Db 1380 GGGGCAAGCCAGAGCGGTGAGCGGACAGCCAGCCAGCCAGCAAAAGCTGTGGGAAGGTG 1439
Qy |||||
Db 1441 GAGCTCCCGGCTCGCCCTGTGCGCCAGTGGAGAGTCCAGGAGCCCAACACCGGTGTCCCC 1500
Qy |||||
Db 1440 GAGCTCCCGGCTCGCCCTGTGCGCCAGTGGAGAGTCCAGGAGCCCAACACCGGTGTCCCC 1499
Qy |||||
Db 1501 GTAGCCAGCCCGCCAAAAGCTGGGATCCAAAACAGCTAAATGCGCACCCGCTGGAGCCT 1560
Qy |||||
Db 1500 GTAGCCAGCCCGCCAAAAGCTGGGATCCAAAACAGCTAAATGCGCACCCGCTGGAGCCT 1559
Qy |||||
Db 1561 GTCTGGGGGCTGAGACAGAGAGGGGCGCTGGCTCGCGCCCGGAGGAAAGGT 1620
Qy |||||

Db 1560 GTCTGGGAGCCTGAAGACAGAGAGGGGCGCTGGCTCGCGCCCGGAGGAAGGT 1619
Qy |||||
Db 1621 TCA 1623
Qy |||||
Db 1620 TCA 1622
Qy |||||
RESULT 7
US-10-615-659-11
; Sequence 11, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: TYROSINE-LIGASE ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-11
Query Match 90.9%; Score 1475.8; DB 19; Length 3465;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1554; Conservative 0; Mismatches 2; Indels 67; Gaps 1;
Qy 1 ATGGCATCCAGCATCTCAAGTGGGTGGTTCAGCCACAGAGTGCAGAGGAGCAGAGCAGAGA 60
Db 1678 ATGGCATCCAGCATCTCAAGTGGGTGGTTCAGCCACAGAGTGCAGAGGAGCAGAGCAGAGA 1737
Qy 61 AGCAAGCCAGGAGCAGAGGAGGAGGCGGGAGAGCGGAGCAGGACCTGAGCAGCAGGACAGAT 120
Db 1738 AGCAAGCCAGGAGCAGAGGAGGAGGCGGGAGAGCGGAGCAGGACCTGAGCAGCAGGCAAGAT 1797
Qy 121 GCTGAATAATGCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC 180
Db 1798 GCTGAATAATGCTGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC 1857
Qy 181 AAGGTGTGCAGGCTTACCTGAGGAGCTGAGGAGCATGAGGACATCGACAGCTCAGCAGAT 240
Db 1858 AAGGTGTGCAGGCTTACCTGAGGAGCTGAGGAGCATGAGGACATCGACAGCTCAGCAGAT 1917
Qy 241 GCGGTGAGGAGCCTCACTGAGGCGAGTGGGAGGAGCTGACCCAGCAGTACTACTCCCTC 300
Db 1918 GCGGTGAGGAGCCTCACTGAGGCGAGTGGGAGGAGCTGACCCAGCAGTACTACTCCCTC 1977
Qy 301 GTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGTGCAGGCTCTG 360
Db 1978 GTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGTGCAGGCTCTG 2037
Qy 361 CTGAATAGAAATCAGTCTGTGAAACCTCAGAGCGGACATTTGACGGGTCCGGAACTCTGG 420
Db 2038 CTGAATAGAAATCAGTCTGTGAAACCTCAGAGCGGACATTTGACGGGTCCGGAACTCTGG 2097
Qy 421 ATTATTAAGCCCGCCGAGTCCCGGGGCGGAGACATAGTGTGCATGGACCGTGTGGAG 480
Db 2098 ATTATTAAGCCCGCCGAGTCCCGGGGCGGAGTCCCGGGGCGG----- 2129
Qy 481 GAGATCTCGAGTGCAGCTGCAGACCACTCTTTTCCAGGGAGCAACAAGTGGGTGGTC 540
Db 2130 -----AGGCAACAAGTGGGTGGTC 2150
Qy 541 CAGAAGTACATCGAGACGCCGCTGTCTCATCTGTGACCAACAAGTTTCGACATCAGACAGTGG 600
Db 2151 CAGAAGTACATCGAGACGCCGCTGTCTCATCTGTGACCAACAAGTTTCGACATCAGACAGTGG 2210
Qy 601 TTCCTCGTCAAGGACTGGAAACCCCTTGACCATCTGTTCTTACAAGGAGAGTTACTTGGCGG 660


```
QY 721 GCCGTCAGAGTACCTGAAGATGATGTGGCGCAGACCCCTGCTGCGCGCACACAAC 780
Db |||||
QY 2331 GCCGTCAGAGTACCTGAAGATGATGTGGCGCAGACCCCTGCTGCGCGCACACAAC 2390
Db |||||
QY 781 ATGTGGACCAAGCAGGTTCCAGGAGTACTGTGAGCGCCAGGCGCGTGGCGCGTGTGG 840
Db |||||
QY 2391 ATGTGGACCAAGCAGGTTCCAGGAGTACTGTGAGCGCCAGGCGCGTGGCGCGTGTGG 2450
Db |||||
QY 841 GGCAGCGTCTATCTACCGCTCATGAAGAGGCGCATTCGCCACGACCAATGAAGGTGGCCAG 900
Db |||||
QY 2451 GGCAGCGTCTATCTACCGCTCATGAAGAGGCGCATTCGCCACGACCAATGAAGGTGGCCAG 2510
Db |||||
QY 901 GACACGTGAGCGCTCGCAAGACAGCTTTGAGCTCTACCGGGCTGACTTCGTCTTGG 960
Db |||||
QY 2511 GACACGTGAGCGCTCGCAAGACAGCTTTGAGCTCTACCGGGCTGACTTCGTCTTGG 2570
Db |||||
QY 961 AGGAGCTTCAAGCGCTGCTGATCGAGATCAATTTCCAGCGCCCAACCATGCAACCGTCCACG 1020
Db |||||
QY 2571 AGGAGCTTCAAGCGCTGCTGATCGAGATCAATTTCCAGCGCCCAACCATGCAACCGTCCACG 2630
Db |||||
QY 1021 CCGGTCAAGCGCCAGCTGTGTGACAGGTGCAAGAGGACACCATCAAGGTGGCGGTGGAC 1080
Db |||||
QY 2631 CCGGTCAAGCGCCAGCTGTGTGACAGGTGCAAGAGGACACCATCAAGGTGGCGGTGGAC 2690
Db |||||
QY 1081 CGCAGCTGTGACATCGCAACTTTCGAGCTCTGTGAGGCGAGCGGTGTGTGAGCGGCC 1140
Db |||||
QY 2691 CGCAGCTGTGACATCGCAACTTTCGAGCTCTGTGAGGCGAGCGGTGTGTGAGCGGCC 2750
Db |||||
QY 1141 CCATTCAAGCGGTTCGCACTCTCGTGGCGGGCTCAGTGTGAGGAGAGCAGGAGGCGAG 1200
Db |||||
QY 2751 CCATTCAAGCGGTTCGCACTCTCGTGGCGGGCTCAGTGTGAGGAGAGCAGGAGGCGAG 2810
Db |||||
QY 1201 GTGCTGCCGTCTGCAACTCAAGGCTTCGGCTCTGCTGTGACGCGCAGCGCTGAAG 1260
Db |||||
QY 2811 GTGCTGCCGTCTGCAACTCAAGGCTTCGGCTCTGCTGTGACGCGCAGCGCTGAAG 2870
Db |||||
QY 1261 GCACGGGGCCCTCGGCCATGCGAGCCCTGCGCCAGGAGCCCATCACAGCTCTCCAG 1320
Db |||||
QY 2871 GCACGGGGCCCTCGGCCATGCGAGCCCTGCGCCAGGAGCCCATCACAGCTCTCCAG 2930
Db |||||
QY 1321 CGGAGCTTGGGACTGAAGAGAGAGAGGGCTCCCTCGCTTGTGGCACCTTTAAGG 1380
Db |||||
QY 2931 CGGAGCTTGGGACTGAAGAGAGAGAGGGCTCCCTCGCTTGTGGCACCTTTAAGG 2990
Db |||||
QY 1381 GGGGCAAGCCAGAGCGGTGGAGCGCAGCAGCCCAACCGCAACCAAGCTGTGGGAAGGTG 1440
Db |||||
QY 2991 GGGGCAAGCCAGAGCGGTGGAGCGCAGCAGCCCAACCGCAACCAAGCTGTGGGAAGGTG 3050
Db |||||
QY 1441 GAGCTCCGCGCTGCGCTGTGCGCAGTGGACAGTTCAGGCGCCCAACACCGGTGTCCCG 1500
Db |||||
QY 3051 GAGCTCCGCGCTGCGCTGTGCGCAGTGGACAGTTCAGGCGCCCAACACCGGTGTCCCG 3110
Db |||||
QY 1501 GTAGCCAGCCCGCAAAAGCTGGATCCAAAACAGCTAAATGGCACCGCTGGAGCCT 1560
Db |||||
QY 3111 GTAGCCAGCCCGCAAAAGCTGGATCCAAAACAGCTAAATGGCACCGCTGGAGCCT 3170
Db |||||
QY 1561 GTGCTGGGGCTGTAAGACAGCAGAGGGCGCGCTGCTCGCGCGCGCGAGGAAAGGT 1620
Db |||||
QY 3171 GTGCTGGGAGCTGAAGACAGCAGAGGGCGCGCTGCTCGCGCGCGCGAGGAAAGGT 3230
Db |||||
QY 1621 TCA 1623
Db |||
QY 3231 TCA 3233
Db |||
```

RESULT 9

```
US-10-615-659-9
; Sequence 9, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
```

```
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
```

```
; FILE REFERENCE: D0283 NP
```

```
; CURRENT APPLICATION NUMBER: US/10/615,659
```

```
; CURRENT FILING DATE: 2003-07-09
```

```
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
```

```
; PRIOR FILING DATE: 2002-07-09
```

```
; NUMBER OF SEQ ID NOS: 102
```

```
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 9
```

```
; LENGTH: 1939
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; US-10-615-659-9
```

```
Query Match 90.8%; Score 1474.2; DB 19; Length 1939;
```

```
Best Local Similarity 95.7%; Pred. No. 0;
```

```
Matches 1553; Conservative 0; Mismatches 3; Indels 67; Gaps 1;
```

```
QY 1 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGGAGCAGCAGA 60
Db |||||
QY 153 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTGCAGCAGGAGCAGCAGA 212
Db |||||
QY 61 AGCAAGCCCAAGGACCCAGAGGAGGAGGCGCGGAGCAGCGACCTGAGCAGCAGCAGAT 120
Db |||||
QY 213 AGCAAGCCCAAGGACCCAGAGGAGGAGGCGCGGAGCAGCGACCTGAGCAGCAGCAGAT 272
Db |||||
QY 121 GCTGAATAATGCTAGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC 180
Db |||||
QY 273 GCTGAATAATGCTAGAGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC 332
Db |||||
QY 181 AAGGTGTGCGAGCCCTACCTGGGGCAGCTGGAGCATGAGGACATCGACAGTCAAGCAGAT 240
Db |||||
QY 333 AAGGTGTGCGAGCCCTACCTGGGGCAGCTGGAGCATGAGGACATCGACAGTCAAGCAGAT 392
Db |||||
QY 241 GCGGTGAGGACCTCACTGAGGCCGAGTGGAGGACCTTGACCCAGCAGTACTACTCCCTC 300
Db |||||
QY 393 GCGGTGAGGACCTCACTGAGGCCGAGTGGAGGACCTTGACCCAGCAGTACTACTCCCTC 452
Db |||||
QY 301 GTTCATGGGATGCTTTTCATCTCCAAATTCAAGAAATTTACTTTTCGAGTGCAGGCTCTG 360
Db |||||
QY 453 GTTCATGGGATGCTTTTCATCTCCAAATTCAAGAAATTTACTTTTCGAGTGCAGGCTCTG 512
Db |||||
QY 361 CTGATAGAGTACGCTGTGAACCTCAGAGCGGACATTCAGCGGCTCCGGAACATCTGG 420
Db |||||
QY 513 CTGAATAGAGTACGCTGTGAACCTCAGAGCGGACATTCAGCGGCTCCGGAACATCTGG 572
Db |||||
QY 421 ATTATTAAGCCCGCGCCAAAGTCCCGGGCCGAGACATAGTGTGATGGACCGTGTGGAG 480
Db |||||
QY 573 ATTATTAAGCCCGCGCCAAAGTCCCGGGCCG----- 604
Db |||||
QY 481 GAGATCCTGAGCTGGCAGCTGCAGACCACTCTTTTCAGGGGACAAAGTGGGTGGTC 540
Db |||||
QY 605 -----AGGACAAACAAGTGGGTGGTC 625
Db |||||
QY 541 CAGAGTATCATCAGACGCGCTGCTCATCTGTGACACCAAGTTCCGACATCAGCAGTGG 600
Db |||||
QY 626 CAGAGTATCATCAGACGCGCTGCTCATCTGTGACACCAAGTTCCGACATCAGCAGTGG 685
Db |||||
QY 601 TTCTCTGTCAGCAGCTGGAAACCCCTGACCATCTGGTTCTACAAGGAGAGTTACTTTGCGG 660
Db |||||
QY 686 TTCTCTGTCAGCAGCTGGAAACCCCTGACCATCTGGTTCTACAAGGAGAGTTACTTTGCGG 745
Db |||||
QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAGTGGACAGCGCCATCCACCTGTGCAACAAC 720
Db |||||
QY 746 TTCTCAACTCAGCGCTTCTCCCTGGACAAGTGGACAGCGCCATCCACCTGTGCAACAAC 805
Db |||||
QY 721 GCGGTCCAGAGTACTCTGAAGATGATGTGGCGCAGCGCCCTGCTGCGCGCACACAAC 780
Db |||||
QY 806 ACCGTCCAGAGTACTCTGAAGAAATGATGTGGCGCGCAGCGCCCTGCTGCGCGCACACAAC 865
Db |||||
QY 781 ATGTGGACCAAGCAGCAGGTTCCAGGAGTACCTGACGCGCAGGCGCGTGGCGCGGTGTGG 840
Db |||||
QY 866 ATGTGGACCAAGCAGCAGGTTCCAGGAGTACCTGACGCGCAGGCGCGTGGCGCGGTGTGG 925
Db |||||
```



```
QY 841 GGACGCTCATCTACCCGCTCATGAAGAGGCGCATCGCCACGCCATGAAGTGGCCCG 900
Db 926 GGCAGCGTCATCTACCCGCTCATGAAGAGGCGCATCGCCACGCCATGAAGTGGCCCG 985
QY 901 GACACGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGCTCTGGG 960
Db 986 GACACGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGCTCTGGG 1045
QY 961 AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGAGCCACCAATGACCGCTCCAG 1020
Db 1046 AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGAGCCACCAATGACCGCTCCAG 1105
QY 1021 CCGGTCA CGGCCCGAGCTGTGTGCACAGGTGCGAGGAGCACCATCAAGGTGGCCGTGGAC 1080
Db 1106 CCGGTCA CGGCCCGAGCTGTGTGCACAGGTGCGAGGAGCACCATCAAGGTGGCCGTGGAC 1165
QY 1081 CGGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGAGGAGCGCGTGTGAGCGCCGCC 1140
Db 1166 CGGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGAGGAGCGCGTGTGAGCGCCGCC 1225
QY 1141 CATTCAGCGGGTCCGACCTCTGCGTGGCGGGCTCAAGTGTGAGGAGCGAGGAGCGAG 1200
Db 1226 CATTCAGCGGGTCCGACCTCTGCGTGGCGGGCTCAAGTGTGAGGAGCGAGGAGCGAG 1285
QY 1201 GTGCTGCCGCTCTGCAACCTCAAGGCTCGGCTCGCTGTGTGACCGCGAGCGCTGAAG 1260
Db 1286 GTGCTGCCGCTCTGCAACCTCAAGGCTCGGCTCGCTGTGTGACCGCGAGCGCTGAAG 1345
QY 1261 GCACGGGGCCCTCGGCGCATGCGAGACCTGCGCCAGGAGCCGCCATCAACAGCTTCCAG 1320
Db 1346 GCACGGGGCCCTCGGCGCATGCGAGACCTGCGCCAGGAGCCGCCATCAACAGCTTCCAG 1405
QY 1321 CGGGACTCGGACTGAAGAGAGAGGGGCTCCCTCGCTGCTGCGCACCTTAAGG 1380
Db 1406 CGGGACTCGGACTGAAGAGAGAGGGGCTCCCTCGCTGCTGCGCACCTTAAGG 1465
QY 1381 GGGGACGCGAGAGCGGTGGAGCCGACAGCCACCCGCGACCAAGCTGCTGGGAAGGTG 1440
Db 1466 GGGGACGCGAGAGCGGTGGAGCCGACAGCCACCCGCGACCAAGCTGCTGGGAAGGTG 1525
QY 1441 GAGCTCCGGCTCGCCCTGTGCGCAAGTGGAGAGTCAAGGCGCCCAACACCGGTGTCGC 1500
Db 1526 GAGCTCCGGCTCGCCCTGTGCGCAAGTGGAGAGTCAAGGCGCCCAACACCGGTGTCGC 1585
QY 1501 GTAGCCCGCCGCAAAAGCTGGATCCAAACCAAGCTAAATGCGCACCCGCTGGAGCCT 1560
Db 1586 GTAGCCCGCCGCAAAAGCTGGATCCAAACCAAGCTAAATGAGCACCCGCTGGAGCCT 1645
QY 1561 GTGCTCGGGGCTGAAGACAGCAGAGGGCGCGCTCGCTCGCGCGCCCGGAGGAAAGGT 1620
Db 1646 GTGCTCGGGGCTGAAGACAGCAGAGGGCGCGCTCGCTCGCGCGCCCGGAGGAAAGGT 1705
QY 1621 TCA 1623
Db 1706 TCA 1708
```

RESULT 10

```
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
```

```
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match      90.8%; Score 1474.2; DB 19; Length 1939;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1553; Conservative 0; Mismatches 3; Indels 67; Gaps 1;

QY 1 ATGGCATCCAGCATCTCAAGTGGGTGTCAGCCACAGAGCTGCAGAGAGGAGCAGACA 60
Db 153 ATGGCATCCAGCATCTCAAGTGGGTGTCAGCCACAGAGCTGCAGAGAGGAGCAGACA 212
QY 61 ACRAAGCCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 120
Db 213 AGCAAGCCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 272
QY 121 GCTGAAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGACATCGCGTGC 180
Db 273 GCTGAAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGACATCGCGTGC 332
QY 181 AAGGTGTGCGAGGCTTACCTGCGGCGAGCTGAGAGCATGAGGACATCGACACGTCAGAGAT 240
Db 333 AAGGTGTGCGAGGCTTACCTGCGGCGAGCTGAGAGCATGAGGACATCGACACGTCAGAGAT 392
QY 241 GCGGTGGAGGACCTCACTGAGGCGCGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 300
Db 393 GCGGTGGAGGACCTCACTGAGGCGCGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAT 452
QY 301 GTTCATGGGATGCTTTCATCTCCAAATTCAGAAATTAATCTTTTCGAGTGCCAGGCTCTG 360
Db 453 GTTCATGGGATGCTTTCATCTCCAAATTCAGAAATTAATCTTTTCGAGTGCCAGGCTCTG 512
QY 361 CTGAATAGAAATCACGCTGTGAAACCTCAGACGGACATTTGACGGGTCCGGAAACATCTGG 420
Db 513 CTGAATAGAAATCACGCTGTGAAACCTCAGACGGACATTTGACGGGTCCGGAAACATCTGG 572
QY 421 ATTATAAGCCCGCGGCGCAAGTCCCGGGGCGAGACATAGTGTGCATGACCGGTGTGGAG 480
Db 573 ATTATAAGCCCGCGGCGCAAGTCCCGGGGCGG----- 604
QY 481 GAGATCCTGAGCTGCGAGCTGACAGACCAACCTCTTTCCAGGAGCAACAAAGTGGGTGGTC 540
Db 605 -----AGGACAAACAAGTGGGTGGTC 625
QY 541 CAGAAATGATATCGAGACGCGCTGTCTATCTGTGACACCAAGTTGACATCGACAGTGG 600
Db 626 CAGAAATGATATCGAGACGCGCTGTCTATCTGTGACACCAAGTTGACATCGACAGTGG 685
QY 601 TTCTCTCGTCACGAGCTGGAAACCCCTGACCATCTGGTCTTACAGGAGAGTACTTTGCCG 660
Db 686 TTCTCTCGTCACGAGCTGGAAACCCCTGACCATCTGGTCTTACAGGAGAGTACTTTGCCG 745
QY 661 TTCTCAACTCAGCGCTTCTCCTTGGAACAAGCTGGACAGCGCCATCCACTGTGCAACAAC 720
Db 746 TTCTCAACTCAGCGCTTCTCCTTGGAACAAGCTGGACAGCGCCATCCACTGTGCAACAAC 805
QY 721 GCGCTCCAGAAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
Db 806 ACCGTCCAGAAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 865
QY 781 ATGTGGACGAGCAGGAGTTCAGGAGTACCTGACGCGGAGGAGGAGGAGGAGGAGGAGGAG 840
Db 866 ATGTGGACGAGCAGGAGTTCAGGAGTACCTGACGCGGAGGAGGAGGAGGAGGAGGAGGAG 925
QY 841 GGCAGGCTCATCTACCCGCTCCATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
Db 926 GGCAGGCTCATCTACCCGCTCCATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 985
QY 901 GACCAGGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGCTCTGGG 960
```


Db 986 GACCACTGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGTGACTTGGTCTTGGG 1045
Qy 961 AGGACTTCAGGCGCTCGGCTGATCGAGATCAATTCAGACCCCAACATGCAACCGTCCAGC 1020
Db 1046 AGGACTTCAGGCGCTCGGCTGATCGAGATCAATTCAGACCCCAACATGCAACCGTCCAGC 1105
Qy 1021 CCGGTACGCGCCAGCTGTGTGCAAGGTGCGAGGAGCACCACCAAGGTGGCGGTGGAC 1080
Db 1106 CCGGTACGCGCCAGCTGTGTGCAAGGTGCGAGGAGCACCACCAAGGTGGCGGTGGAC 1165
Qy 1081 CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGAGGAGCAGCGGTGGTTGAGCCGCC 1140
Db 1166 CGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGAGGAGCAGCGGTGGTTGAGCCGCC 1225
Qy 1141 CCATTACGCGGTCCGACCTCTGCGTGGCGGGCGTCAAGTGTGAGGAGCAGCAGAGAGGAG 1200
Db 1226 CCATTACGCGGTCCGACCTCTGCGTGGCGGGCGTCAAGTGTGAGGAGCAGCAGAGAGGAG 1285
Qy 1201 GTGCTGCCCGCTCTGCAACCTCAAGGCTTCGAGCTCTGAGGAGCAGCGGTGGTTGAGCGCGCTGAAG 1260
Db 1286 GTGCTGCCCGCTCTGCAACCTCAAGGCTTCGAGCTCTGAGGAGCAGCGGTGGTTGAGCGCGCTGAAG 1345
Qy 1261 GCAGCGGCGCTCTGCGCATGCGCAGACCTCTGCGCAGGAGCAGCGGTGGTTGAGCGCGCTGAAG 1320
Db 1346 GCAGCGGCGCTCTGCGCATGCGCAGACCTCTGCGCAGGAGCAGCGGTGGTTGAGCGCGCTGAAG 1405
Qy 1321 CGGACTTTGGGACTGAAGAGAGAGAGGGGTCTCCCTCTGCGCTTGTGGCAACCTTTAAGG 1380
Db 1406 CGGACTTTGGGACTGAAGAGAGAGAGGGGTCTCCCTCTGCGCTTGTGGCAACCTTTAAGG 1465
Qy 1381 GGGGACGCGAGAGCGGTGGAGCGGCGCAGCGCAGCGCAGCGGCTGGAGGAGGTG 1440
Db 1466 GGGGACGCGAGAGCGGTGGAGCGGCGCAGCGCAGCGGCTGGAGGAGGTG 1525
Qy 1441 GAGCTCCCGGCTCTGCGCAGCTGGAAGAGAGGGGTCTCCCTCTGCGCTTGTGGCAACCTTTAAGG 1500
Db 1526 GAGCTCCCGGCTCTGCGCAGCTGGAAGAGAGGGGTCTCCCTCTGCGCTTGTGGCAACCTTTAAGG 1585
Qy 1501 GTAGCCAGCGCCGCAAAAGCTGGATCCAAACAGCTAAATCGCACCCGCTGGAGCCT 1560
Db 1586 GTAGCCAGCGCCGCAAAAGCTGGATCCAAACAGCTAAATCGCACCCGCTGGAGCCT 1645
Qy 1561 GTGCTGGGCGCTTGAAGACAGCAGAGGCGCGCTGCGTCCGCGCGCGGAGGAAAGGT 1620
Db 1646 GTGCTGGGCGCTTGAAGACAGCAGAGGCGCGCTGCGTCCGCGCGCGGAGGAAAGGT 1705
Qy 1621 TCA 1623
Db 1706 TCA 1708

RESULT 11

US-10-615-659-3
; Sequence 3, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS

; LOCATION: (1)..(726)
US-10-615-659-3

Query Match 44.7%; Score 726; DB 19; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.4e-189;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 397 ATTGACGGGCTCGGAAACATCTGGATTATAAGCCCGCGGCAAGTCCCGGGGCGGAGAC 456
Db 1 ATTGACGGGCTCGGAAACATCTGGATTATAAGCCCGCGGCAAGTCCCGGGGCGGAGAC 60
Qy 457 ATAGTGTGCATGACACCGCTGTGGAGGAGATCTTGAGCTGGCAGCTGCAGACCACTCTT 516
Db 61 ATAGTGTGCATGACACCGCTGTGGAGGAGATCTTGAGCTGGCAGCTGCAGACCACTCTT 120
Qy 517 TCCAGGGAACAAGTGGGTCCAGAAATGATCATCGAGACGCGCTGCTCATCTGTGAC 576
Db 121 TCCAGGGAACAAGTGGGTCCAGAAATGATCATCGAGACGCGCTGCTCATCTGTGAC 180
Qy 577 ACCAAGTTCGACATCAGACAGTGGTTCCTGTCACGCACTGGACCCCTGACCATCTGG 636
Db 181 ACCAAGTTCGACATCAGACAGTGGTTCCTGTCACGCACTGGACCCCTGACCATCTGG 240
Qy 637 TTCTACAAGGAGAGTACTTTCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGAC 696
Db 241 TTCTACAAGGAGAGTACTTTCGCTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGAC 300
Qy 697 AGCGCCATCCACTGTGTGCAACAACGCGCTCCAGAAATGATGTGGGCGC 756
Db 301 AGCGCCATCCACTGTGTGCAACAACGCGCTCCAGAAATGATGTGGGCGC 360
Qy 757 AGCGCCCTGCTGCGGAGAGTGGTTCCTGTCACGCACTGGACCCCTGACCATCTGG 816
Db 361 AGCGCCCTGCTGCGGAGAGTGGTTCCTGTCACGCACTGGACCCCTGACCATCTGG 420
Qy 817 CGCAGGCGCTGCGGCGCTGTCGAGGAGTTCATCTACCGCTCCATGAAAGAGCCATC 876
Db 421 CGCAGGCGCTGCGGCGCTGTCGAGGAGTTCATCTACCGCTCCATGAAAGAGCCATC 480
Qy 877 GCCCAGCCCATGAAGTGGCCAGGACCACTGAGGCTCCGCAAGAACAGCTTTGAGCTC 936
Db 481 GCCCAGCCCATGAAGTGGCCAGGACCACTGAGGCTCCGCAAGAACAGCTTTGAGCTC 540
Qy 937 TAGCGGCTGACTTCTGCTTGGAGGAGTTCAGGCGCTGGCTGATCGAGATCAATTC 996
Db 541 TAGCGGCTGACTTCTGCTTGGAGGAGTTCAGGCGCTGGCTGATCGAGATCAATTC 600
Qy 997 AGCCCCACCATGACACCGCTCCAGCGCTCAGCGCCAGCTGTGTGCAAGGTGACAGG 1056
Db 601 AGCCCCACCATGACACCGCTCCAGCGCTCAGCGCCAGCTGTGTGCAAGGTGACAGG 660
Qy 1057 GACACCATCAAGTGGCGCTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 1116
Db 661 GACACCATCAAGTGGCGCTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCTGTGG 720
Qy 1117 AGGCAG 1122
Db 721 AGGCAG 726

RESULT 12

US-10-635-977-3
; Sequence 3, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09


```
; PRIORITY APPLICATION NUMBER: U.S.10/615.659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-635-977-3

Query Match      44.7%; Score 726; DB 19; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.4e-189;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 ATTGACGGGCTCCGGAACATCTGGATTATTAAGACCCCGCGGCCAAGTCCCGGGGCCGAGAC 456
DB 1 ATTGACGGGCTCCGGAACATCTGGATTATTAAGACCCCGCGGCCAAGTCCCGGGGCCGAGAC 60

QY 457 ATAGTGTGATGACGACCGTGTGGAGGAGATCCTGGAGCTGSCAGCTGCAGACCAACCTCTT 516
DB 61 ATAGTGTGATGACCGTGTGGAGGAGATCCTGGAGCTGSCAGCTGCAGACCAACCTCTT 120

QY 517 TCCAGGACAAAGTGGGTGGTCCAGAAATATATCGAGACGCGCTGCTCATCTGTGAC 576
DB 121 TCCAGGACAAAGTGGGTGGTCCAGAAATATATCGAGACGCGCTGCTCATCTGTGAC 180

QY 577 ACCAAGTTGACATCAGACAGATGGTTCCTCGTCAAGGACTGGAACCCCTGACCACTGG 636
DB 181 ACCAAGTTGACATCAGACAGATGGTTCCTCGTCAAGGACTGGAACCCCTGACCACTGG 240

QY 637 TTCTCAAGGAGATTTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 696
DB 241 TTCTCAAGGAGATTTACTTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGAC 300

QY 697 AGCGCCATCCACTGTGCAACACGCGTCCAGAAATACCTGAAGAATGATGTGGGCCG 756
DB 301 AGCGCCATCCACTGTGCAACACGCGTCCAGAAATACCTGAAGAATGATGTGGGCCG 360

QY 757 AGCCCCCTGTCGCCGACACAAATGTGGAACAGACAGAGTTCAGGAGTACCTGCGAG 816
DB 361 AGCCCCCTGTCGCCGACACAAATGTGGAACAGACAGAGTTCAGGAGTACCTGCGAG 420

QY 817 CGCCAGGGCGGTGGCGCGTGTGGGCGAGCGTTCATCTACCGTCCATGAAGAGGCCATC 876
DB 421 CGCCAGGGCGGTGGCGCGTGTGGGCGAGCGTTCATCTACCGTCCATGAAGAGGCCATC 480

QY 877 GCCCAGCCCATGAAGTGGCCAGGACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCTC 936
DB 481 GCCCAGCCCATGAAGTGGCCAGGACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCTC 540

QY 937 TAGGGGCTGACTTCGTCTCTGGAGGGAATTTCAGGCCCTGCTGATCGAGATCAATTC 996
DB 541 TAGGGGCTGACTTCGTCTCTGGAGGGAATTTCAGGCCCTGCTGATCGAGATCAATTC 600

QY 997 AGCCCCACCATGACCCCTGCAAGCGGTCAGCGCCAGCTGTGTCACAGGTGACAGGAG 1056
DB 601 AGCCCCACCATGACCCCTGCAAGCGGTCAGCGCCAGCTGTGTCACAGGTGACAGGAG 660

QY 1057 GACACCATCAAGTGGCCGCTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 1116
DB 661 GACACCATCAAGTGGCCGCTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGG 720

QY 1117 AGGCAG 1122
DB 721 AGGCAG 726

RESULT 13
US-10-723-860-631/c
; Sequence 631, Application US/10723860

; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 631
; LENGTH: 101270
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-631

Query Match      30.9%; Score 501; DB 20; Length 101270;
Best Local Similarity 100.0%; Pred. No. 2.9e-127;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1120 CAGCCGGTGGTTAGCGCGCCCATTCAGCGGTCGCACTCTGCGTGGCGGCGTCAGT 1179
DB 43637 CAGCCGGTGGTTAGCGCGCCCATTCAGCGGTCGCACTCTGCGTGGCGGCGTCAGT 43578

QY 1180 GTGAGGAGCCAGGAGCGAGTGTGCCGCTGCGCACTCAAGCCCTCGGCTCGCTG 1239
DB 43577 GTGAGGAGCCAGGAGCGAGTGTGCCGCTGCGCACTCAAGCCCTCGGCTCGCTG 43518

QY 1240 TTGGAGCGCGACGCTGAAGCAAGCGGCGCCCTCGGCCATGCCAGACCTTGCCAGGGA 1299
DB 43517 TTGGAGCGCGACGCTGAAGCAAGCGGCGCCCTCGGCCATGCCAGACCTTGCCAGGGA 43458

QY 1300 CCCCACATCAAGTCTCCAGCGGACTTTGGGACTGAAGAAAGAGAGGGGCTCCCCCTG 1359
DB 43457 CCCCACATCAAGTCTCCAGCGGACTTTGGGACTGAAGAAAGAGAGGGGCTCCCCCTG 43398

QY 1360 GCCTTGTGCGACCTTAAGGGGGCGAGCGGTGGAGCGCGCACAGCCACCCCGC 1419
DB 43397 GCCTTGTGCGACCTTAAGGGGGCGAGCGGTGGAGCGCGCACAGCCACCCCGC 43338

QY 1420 ACCAAAGCTGTGGGAAGTGGAGCTCCCGGCTGCGCCCTGCGCCACGCTGGACACTCAG 1479
DB 43337 ACCAAAGCTGTGGGAAGTGGAGCTCCCGGCTGCGCCCTGCGCCACGCTGGACACTCAG 43278

QY 1480 GCCCAAAACACCGTGTCCCGTAGCCAGCCCGCCAAAGCTGGGATCCAAACCAAGCTA 1539
DB 43277 GCCCAAAACACCGTGTCCCGTAGCCAGCCCGCCAAAGCTGGGATCCAAACCAAGCTA 43218

QY 1540 AATGGCACCCGCTGGAGCTGTGTCGGGGCCCTGAACACAGCAGAGGCGCGCTGCGT 1599
DB 43217 AATGGCACCCGCTGGAGCTGTGTCGGGGCCCTGAACACAGCAGAGGCGCGCTGCGT 43158

QY 1600 CCGCCGCCCGGAGGAAAAGGT 1620
DB 43157 CCGCCGCCCGGAGGAAAAGGT 43137

RESULT 14
US-10-275-595A-39
; Sequence 39, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dying Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
```


APPLICANT: LAL, Presti
APPLICANT: YAO, Monique G.
APPLICANT: BANDMAN, Olga
APPLICANT: BURFORD, Neil
APPLICANT: BATRA, Sajeev
APPLICANT: KEARNEY, Liam
APPLICANT: POLICKY, Jennifer L.
TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
FILE REFERENCE: PF-0772 USN
CURRENT APPLICATION NUMBER: US/10/275,595A
PRIORITY FILING DATE: 2003-06-13
PRIORITY FILING DATE: 2000-05-05
PRIORITY FILING DATE: 2000-05-08
PRIORITY FILING DATE: 2000-05-29
PRIORITY FILING DATE: 2000-06-05
PRIORITY FILING DATE: 2000-06-07
PRIORITY FILING DATE: 2000-06-21
PRIORITY FILING DATE: 2000-06-21
NUMBER OF SEQ ID NOS: 68
SOFTWARE: PERL Program
SEQ ID NO 39
LENGTH: 2380
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No: 2156553CB1
US-10-275-595A-39

Query Match 26.7%; Score 432.8; DB 18; Length 2380;

Best Local Similarity 59.1%; Pred. No. 9.1e-109;

Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;

108 CAGCAGCAGATGCTGAAATCTGAGGCAAAAGCTCAG-GGGCTCCCGGGCAGCTTG 166
361 CAGGAGCAACAGAGCCCAAGAAACAGAGAAACCCAGAGTGGTGTCCCGCAGAGTTTG 420
167 TGGACATCGCGTGCAAGGTGTGCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCG 226
421 TGGATGAGCTCTGTGTGCTGCGAGGAGTACCTTAGCAACTTGGCCCATGACATGACATCG 480
227 ACACGTGACAGATGCGGTGGAGGACCTCTACAGGCGGAGTGGAGGACCTGACATCAGC 286
481 ACAAGGACCTGGAGGCGCCCTGTACTACCCCGAGGCGTGGTCCCTTCTCTCAGC 540
287 AGTACTACTCCCTGCTTCATGGCGATGCTTTCATCTCCANTTCAAGAAATTACTTTTCG 346
541 GCTACTACCAAGTGGTCCAGAAAGGGGCGAGAACTCAGGACCTCGACATCAGGTCCAGC 600
347 AGTGCCAGGCTCTGCTGAATAGAAATCAGTCTGTGAACCTCAGACGACATTTGACGGGC 406
601 GCTGTGAGGACATCTGCGACAGCTGCAGGCGGTGGTACCCAGATAGACATGGAAGGGG 660
407 TCCGGAACTCTGGATTTATAAGCCCGCCGCAAGTCCCGGGCCGAGACATAGTGTGCA 466
661 ATGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGTGGACGAGCATCATGTGCA 720
467 TGAACCGTGTGGAGGAGATCTGAGGCTGCGACCTGACAGACCACTCTTTCAGGGACA 526
721 TGGACCACTTGGAGGAGATGCTGAAGCTGGTGAACCGGCAACCCCGTGGTGAAGGAGC 780
527 ACAAGTGGGTGGTCCAGAACTACATCAGACGCGCTGCTCATCTGTGACCACTGATTCG 586
781 GCAAGTGGGTGGTGCAGAACTATATTGAGGCGGCCCTCTCATCTTTGGCACCAGTTTG 840
587 ACATCAGACAGTGGTTCTCGTCCAGGACTGGAAACCCCTGACCATCTGTTTCTACAAG 646
841 ACCTCAGACAGTGGTTCTCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 900
647 AGAGTTACTTGGCGGTCTCAACTCAGGCGTTCCTCCTGGACAAAGCTGGACAGCGCCATCC 706

RESULT 15

US-10-029-386-24894/c

; Sequence 24894, Application US/10029386

; Publication No. US20030194704A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI

; FILE REFERENCE: ABOMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029,386

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 34288

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 24894

; LENGTH: 490

901 ACAGCTATATCCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAACCCTGGACACTCAGTGC 960
707 ACCTGTGCAACACAGCGCTCCAGAAGTACCTGAAGAATGATGTGGCCCGCAGCCCCCTGC 766
961 ACCTGTGCAACAACTCCATCCAGAAGCACCTGGAGAATCATGCCATCGGCATCCACTGC 1020
767 TGGCCGACACACATGTTGGACACGACACAGGTTCCAGGAGTACCTGCAGCGCCAGGGCC 826
1021 TTCCGCGCAGACAAATGTTGTTCTAGCCAGAGGTTCCAGGCCCCACCTTGCAGGAGATGGGT 1080
827 GTGGCCCGTGTGGGCGAGGTCATCTACCCGTCATGAAGAAGGCCATCGCCACACCCA 886
1081 CCCCAAATGCTTGGTCCACCATCATCTGCTGGCATGAAGATGCTGTGATCCAGGCAC 1140
887 TGAAGGTGGCCACAGACACGTCGAGCCTTGCAAGAACAGCTTTTGAGCTCTACGGGGCTG 946
1141 TTCAGACCTCCACAGGACACCGTCAGTGTGGAAGGCCAGCTTTGAGCTCTATGGGGCTG 1200
947 ACTTCGTCTTGGAGGGAGCTTTCAGGCGCTTGGCTGTGATCGAGATCAATTCAGGCCCCACCA 1006
1201 ACTTCGTGTTCGGGGAGGAGCTTTCAGGCGCTTGGCTGTGATCGAGATCAACGCGCCACCA 1260
1007 TGCACCGCTCCAGCGCGTCAGGCGCCAGCTGTGTCACAGGTGTCAGGAGGACACCATCA 1066
1261 TGGCACCTCCACAGCAGTCACTGCCCCGGCTCTGTGCTGGCGTGAAGCTGACACCTGTC 1320
1067 AGGTGGCC-----GTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGT 1114
1321 GGTGTGCTATTGACCGGATGCTGGACCGCACTGTGACACAGGAGCTTTGAGCTCATCT 1380
1115 GAGGCGAGCGGTGTGAGCGCGCCCAATTCAGCGGCTCGACCTTGGCTGGCGGGCG 1174
1381 ATAAGCAGCTGTGTGTGGAGTGCCTCAATATGTGGGCATCCGGCTCTGTTAGAGGGCT 1440
1175 TCAGTGTGAGGAGAGCAGGAGGAGTGTGCTGCCCTGTCACACTCAAGGCTCGGCT 1234
1441 TCACATCAAGAACCCCATGGCGATGTGTCATCGGCGGATGGGGGTCCGCCAGCAGTCC 1500
1235 CGCTGTGACGCGCAGCGCTGAAGGCACGGGGCCCTCTGGCCATGCC-AGACCTTGC 1293
1501 CTCTGTGACCCAGCAGGCTCTGGGGAAGGCAAGGACTCGGGGATCCCTACCCACAGT 1560
1294 CAGGAGCCCCATCACCAGCTCTCAGCGGAGCTTGGAGTGAAGAGAGAGGGGCTC 1353
1561 CAGCTTCTAGGAAGGCACTGGGGCCAGGAGCTGGGGCCACAGTGAAGAGCCAGTCTCCA 1620
1354 CCCTGGCTTGTGTCACCTTAAAGGGGGGAGCGAGAGCGGTGGAGCGGCACAGCCC 1413
1621 CTGCCACCACTTCAGCCCCCGGAAAGGGAAGAAAGGCAAGGCGAAAGGGCCACAGCCC 1680
1414 ACCCGCACAAAGCTGCTGGGAAGGTGGAGCTCCCGGC 1451
1681 TGGTCTGCCCAATCTCTGGGAGTGGGATGCCCCAGC 1718

RESULT 15

US-10-029-386-24894/c

; Sequence 24894, Application US/10029386

; Publication No. US20030194704A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI

; FILE REFERENCE: ABOMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029,386

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 34288

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 24894

; LENGTH: 490


```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022327.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.35
; OTHER INFORMATION: SWISSPROT HIT: P38160, EVALUE 5.00e-04
; OTHER INFORMATION: EST_HUMAN HIT: BB827753.1, EVALUE 3.00e-06
; OTHER INFORMATION: NT HIT: G114726145, EVALUE 8.00e-09
US-10-029-386-24894

Query Match      26.3%; Score 427; DB 16; Length 490;
Best Local Similarity 100.0%; Pred. No. 2.8e-107; Indels 0; Gaps 0;
Matches 427; Conservative 0; Mismatches 0;

QY 696 CAGCGCCATCCACCTGTGCAACAACGCCGTCCAGAAAGTACCTGAAGAATGATGGGCG 755
Db      |||
QY 427 CAGCGCCATCCACCTGTGCAACAACGCCGTCCAGAAAGTACCTGAAGAATGATGGGCG 368
Db      |||
QY 756 CAGCCCCCTGCTGCCCGGCACACACATGTGGACCAGCACAGGTTCCAGGAGTACCTGCA 815
Db      |||
QY 367 CAGCCCCCTGCTGCCCGGCACACACATGTGGACCAGCACAGGTTCCAGGAGTACCTGCA 308
Db      |||
QY 816 GCGCCAGGGCCGTGGGCGCGTGTGGGGCAGCGTCATCTACCCGTCCATGAAGAAGGCCAT 875
Db      |||
QY 307 GCGCCAGGGCCGTGGGCGCGTGTGGGGCAGCGTCATCTACCCGTCCATGAAGAAGGCCAT 248
Db      |||
QY 876 CGCCCCACGCCATGAAGGTGGCCCGCAGGACCAACCGTGGAGCCCTCGCAAGAACAGCTTTGAGCT 935
Db      |||
QY 247 CGCCCCACGCCATGAAGGTGGCCCGCAGGACCAACCGTGGAGCCCTCGCAAGAACAGCTTTGAGCT 188
Db      |||
QY 936 CTACGGGGCTGACTTCGTCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTC 995
Db      |||
QY 187 CTACGGGGCTGACTTCGTCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTC 128
Db      |||
QY 996 CAGCCCCACCATGSCACCCGTCCACGCCCGTCCAGGCCAGCTGTGTGCACAGGTGCAGGA 1055
Db      |||
QY 127 CAGCCCCACCATGSCACCCGTCCACGCCCGTCCAGGCCAGCTGTGTGCACAGGTGCAGGA 68
Db      |||
QY 1056 GGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCCTGTG 1115
Db      |||
QY 67 GGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCCTGTG 8
Db      |||
QY 1116 GAGGCAG 1122
Db      |||
QY 7 GAGGCAG 1
```

Search completed: September 24, 2005, 16:48:40
Job time : 1255.07 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 20:40:04 ; Search time 296.571 Seconds
(without alignments)
8954.609 Million cell updates/sec

Title: US-10-615-659-1_COPY_153_1775
Perfect score: 1623
Sequence: 1 atggcaccagcactctcaa.....cgcccgaggaaaggttca 1623

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	421.4	26.0	3001	4	US-09-620-312D-494
2	56	3.5	909	4	US-09-252-991A-5992
3	56	3.5	1305	4	US-09-252-991A-5951
4	52.6	3.2	1272	4	US-09-252-991A-2471
5	52.6	3.2	1449	4	US-09-252-991A-2200
6	52.6	3.2	2790	4	US-09-252-991A-2632
7	50.4	3.1	1524	4	US-09-807-897-3
8	50.2	3.1	9039	4	US-09-902-540-983
9	48.6	3.0	2277	1	US-08-676-967-5
10	48.6	3.0	2277	1	US-08-676-974-5
11	48.6	3.0	2277	2	US-09-098-487-5
12	48.2	3.0	2004	1	US-08-471-033-18
13	48.2	3.0	2004	2	US-08-471-044-18
14	48.2	3.0	2004	2	US-08-463-483A-18
15	48.2	3.0	2004	2	US-08-471-046A-18
16	48.2	3.0	2004	2	US-08-470-566B-18
17	48.2	3.0	2004	2	US-08-469-334-18
18	48.2	3.0	2004	3	US-09-300-529-18
19	48.2	3.0	2655	1	US-08-471-033-17
20	48.2	3.0	2655	2	US-08-471-044-17
21	48.2	3.0	2655	2	US-08-463-483A-17
22	48.2	3.0	2655	2	US-08-471-046A-17
23	48.2	3.0	2655	2	US-08-470-566B-17
24	48.2	3.0	2655	2	US-08-469-334-17
25	48.2	3.0	2655	3	US-09-300-529-17
26	48	3.0	5314	4	US-09-620-312D-138
27	47.6	2.9	927	4	US-09-902-540-9188

28	47.6	2.9	4884	4	US-09-252-991A-12126	Sequence 12126, A
29	47.6	2.9	4884	4	US-09-252-991A-12292	Sequence 12292, A
30	46.4	2.9	4600	4	US-09-702-705-1797	Sequence 1797, Ap
31	46.4	2.9	4600	4	US-09-736-457-1797	Sequence 1797, Ap
32	46.4	2.9	4600	4	US-09-671-325-1797	Sequence 1147, Ap
33	46.4	2.9	19455	4	US-09-902-540-1147	Sequence 1147, Ap
34	46	2.8	1218	4	US-09-902-540-5379	Sequence 5379, Ap
35	45	2.8	2847	4	US-09-902-540-7588	Sequence 7588, Ap
36	46	2.8	7704	4	US-09-902-540-743	Sequence 743, App
37	46	2.8	34552	4	US-09-902-540-1262	Sequence 1262, Ap
38	45.8	2.8	1515	4	US-09-902-540-9184	Sequence 9184, Ap
39	45.8	2.8	9556	4	US-09-902-540-982	Sequence 982, App
40	45.6	2.8	534	4	US-09-252-991A-7596	Sequence 7596, Ap
41	45.6	2.8	576	4	US-09-252-991A-7886	Sequence 7886, Ap
42	45.6	2.8	1047	4	US-09-252-991A-14678	Sequence 14678, A
43	45.6	2.8	1566	4	US-09-902-540-4578	Sequence 4578, Ap
44	45.6	2.8	1686	4	US-09-252-991A-14548	Sequence 14548, A
45	45.6	2.8	1695	4	US-09-252-991A-7816	Sequence 7816, Ap

ALIGNMENTS

RESULT 1
US-09-620-312D-494
; Sequence 494, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Felyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungling
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt.PL_genes Version 1.0
; SEQ ID NO 494
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1361)..(2419)
US-09-620-312D-494

Query Match 26.0%; Score 421.4; DB 4; Length 3001;
Best Local Similarity 53.8%; Pred. No. 1.6e-89;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 80 GGAGGAGGCGGGAGCAGCGACCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAA 139
|||||
Db 1223 GGATCGTGACAGGCCCTCCCTATGTGAGGAGACAGCAGCCCAAGAACAGGAGAAA 1282

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5951
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5951

Query Match 3.5%; Score 56; DB 4; Length 1305;
Best Local Similarity 44.4%; Pred. No. 0.0014;
Matches 224; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

Qy 780 CATGTGACACAGCACCAGGTTCCAGGAGTACTCTCAGCGCCAGCGCGCTGCGCGCGGTG 839
Db 546 CCGTATGAGCCTCACCGCGCGCGCGGACAGAGAGAGCGCGCGTCCGGTGAAGTTCGG 605
Qy 840 GGGCAGCGGTTCATCTACCGCGTTCATGAAGAAGGCCATCGCCACGCCATGAAGTGGCCCA 899
Db 606 CGTGGCGCTGACCGATATCTCTACCGGGCTGTACTCTCCACCGCGGTGCTCGCGCGCCT 665
Qy 900 GGACACGTGGAGCCTCGAAGAACAGCTTTGAGCTCTACGGGCTGACTTCGTCTCTGG 959
Db 666 CGCCACCGCGACGCTCAGCGGCATCGGCCAGCACATCGACATGGCCCTGCTCGACGTGCA 725
Qy 960 GAGGACTTCAGGCGCTGCTGTCGAGATCAATTCCAGCGCCACCATGACCGTCCAC 1019
Db 726 GGTGCGCTGCTGCGCAACAGACACCTCACTGACCAACCGCGCGTCCGCGCGCGG 785
Qy 1020 GCGCGTCACGCGCCAGCTGTGTGCACAGGTGCGAGGAGACACCATCAAGTGGCGGTGA 1079
Db 786 CTTGGGCAACGACACCGAATCGTGGCTACCGAGCTTCCCAACCGCGAGCGGA 845
Qy 1080 CCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGTGAGGAGCGCGTGTGTGAGCGCC 1139
Db 846 CATGATCTCAGCGTCCGCAACGACAGCCAGTTCGCAAGTTCCGCGAGCTGCGAGACCA 905
Qy 1140 CCCATTACGCGGCTCCGACCTCTCGTGGCGGGGTGAGTGTGAGGAGAGCCAGGACCA 1199
Db 906 CCGGGAATGGCGACACCGCGCTTCGCCACCAAGCGCGGTGGCCAAACCGCGA 965
Qy 1200 GGTGCTGCGCGCTCTGCAACCTCAAGGCGCTCGCTGTGAGCGCGAGCGCTGAA 1259
Db 966 GTGTCTGATCCGCTGATCCGCGAGGCGACGCTGCTGCACACCAACCGCGAGTCTCT 1025
Qy 1260 GGCACGGGCGCCCTCGGCCATGCC 1283
Db 1026 TTCCTGGAGCGCGCGCGTGC 1049

RESULT 4

US-09-252-991A-2471
; Sequence 2471, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2471
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2471

Query Match 3.2%; Score 52.6; DB 4; Length 1272;
Best Local Similarity 48.5%; Pred. No. 0.0088;

Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
Qy 393 GGACATTGACGGGCTCGGAAACATCTGGATTATAAGCCCGCGCCAGTCCCGGGCGG 452
Db 756 GGAATGCTCGCGCAGCGCGGACGAGTTCGGCACCAGCGGAGGACGATCCT 815
Qy 453 AGACATAGTGTGCATGACCGTGTGGAGGAGATCTTGGAGCTGCAGCTGCAGACCAACC 512
Db 816 CGACGTACGATGGTTCGACCGCGGCGCGCTGGAGTGGTTCAGGTAGGCGCAGGC 875
Qy 513 TCTTTCCAGGGACAACAAGTGGGTGCTCCAGAAAGTACATCGAGACGCGCTGTCTCATCTG 572
Db 876 GGTGACGCTGGAAGTTCGAGTTCGAGTTCGCCAGGACATCGAGCGGCTGATCCTCGGCTT 935
Qy 573 TGACACCAAGTTGCATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCAT 632
Db 936 CATGATCAAGGACCGCTCGGCCAGCGATGTACGGGATCAATACCCACCGCTGGACAA 995
Qy 633 CTGGTTCTACAAGGAGATTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 691
Db 996 GCGCTCACCGACTGAAGCCCGCGAGCGCATCACTACCGCTTCGCTTCGACATGC 1054

RESULT 5

US-09-252-991A-2200/c
; Sequence 2200, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2200
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2200

Query Match 3.2%; Score 52.6; DB 4; Length 1449;
Best Local Similarity 48.5%; Pred. No. 0.0092;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

Qy 393 GGACATTGACGGGCTCGGAAACATCTGGATTATAAGCCCGCGCCAGTCCCGGGCGG 452
Db 562 GGAATGCTCGCGCAGCGCGGACGAGTTCGGCACCAGCGGAGGACGATCCT 503
Qy 453 AGACATAGTGTGCATGACCGTGTGGAGGAGATCTTGGAGCTGCAGCTGCAGACCAACC 512
Db 502 CGACGTACGATGGTTCGACCGCGGCGCGCTGGAGTGGTTCAGGTAGGCGCAGGC 443
Qy 513 TCTTTCCAGGGACAACAAGTGGGTGCTCCAGAAAGTACATCGAGACGCGCTGTCTCATCTG 572
Db 442 GGTGACGCTGGAAGTTCGAGTTCGAGTTCGCCAGGACATCGAGCGGCTGATCCTCGGCTT 383
Qy 573 TGACACCAAGTTGCATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCAT 632
Db 382 CATGATCAAGGACCGCTCGGCCAGCGATGTACGGGATCAATACCCACCGCTGGACAA 323
Qy 633 CTGGTTCTACAAGGAGATTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 691
Db 322 GCGCTCACCGACTGAAGCCCGCGAGCGCATCACTACCGCTTCGCTTCGACATGC 264

RESULT 6

US-09-252-991A-2632
; Sequence 2632, Application US/09252991A


```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2632
; LENGTH: 2790
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2632

Query Match      3.2%; Score 52.6; DB 4; Length 2790;
Best Local Similarity 48.5%; Pred. No. 0.011;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 393 GGACATTGCGGCTCGGACATCTGGATTATAAGCCGCGGCGCAAGTCCCGGGCG 452
Db 895 GGAATGCTCGCGAGCGCGGAGTACGGACGATTTCCGGCACCGCGGAGGCGGATCCT 954
QY 453 AGACATAGTGTGCATGCAGCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACC 512
Db 955 CGACGTACGGATGTTGCACCGCGGAGCGCGCTGGAGGTGTGAGTAGGCGACGCG 1014
QY 513 TCTTTCCAGGACAAACAAGTGGTGTCTCAGAAATACATCGAGACGCGCTGCTCATCTG 572
Db 1015 GGTGACGCTGGAAGTCAAGTTCGAGTTCGCCAGGACATCGAGCGCTGATCTCGGCTT 1074
QY 573 TGACACCAAGTTGCATCAGACAGTGTCTCTCGTACAGGACTGAAACCCCTGACCAT 632
Db 1075 CATGATCAAGGACCGCTCGGCCAGCGGATGTACGGGATCAATACCCACCGCTGGACAA 1134
QY 633 CTGGTTCTACAAGGAGATTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC 691
Db 1135 GCGCTCAGCGACTGAAGCCGGCGAGCGCATCACTACCGCTTCGCTTCGACATGC 1193

RESULT 7
US-09-807-897-3
; Sequence 3, Application US/09807897
; Patent No. 6660851
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Yoshimitsu
; APPLICANT: UEKI, Jun
; TITLE OF INVENTION: NOVEL DNA FRAGMENT ENHANCING GENE EXPRESSION LEVEL
; FILE REFERENCE: 0230-0156P
; CURRENT APPLICATION NUMBER: US/09/807,897
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: PCT/JP00/05539
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Oryza sativa L.
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Variety: IR24 - Tissue: pistil
; OTHER INFORMATION: Library name: lambdaZAPII cDNA library derived from pistil mRNA
; OTHER INFORMATION: Clone name: RPC213
; NAME/KEY: misc feature
; OTHER INFORMATION: Strandedness: double - Topology: linear
; OTHER INFORMATION: Molecular type: cDNA to mRNA
; OTHER INFORMATION: Feature: nt1, nt2, nt3: transcription initiation points of RPC213
; OTHER INFORMATION: gene determined by primer extension
```

```
; NAME/KEY: misc feature
; OTHER INFORMATION: nt22-nt24: the first ATG of RPC213 gene
; OTHER INFORMATION: nt134-nt136: the second ATG of RPC213 gene
; OTHER INFORMATION: nt295-nt297: the third ATG of RPC213 gene
; OTHER INFORMATION: nt1276-nt1278: termination codon of RPC213 gene
; NAME/KEY: misc feature
; OTHER INFORMATION: nt1343-nt1348: nt1365-nt1370: poly (A) signals
; OTHER INFORMATION: nt1507-nt1524: poly (A)
US-09-807-897-3

Query Match      3.1%; Score 50.4; DB 4; Length 1524;
Best Local Similarity 47.2%; Pred. No. 0.031;
Matches 187; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 811 CTGCAGCGCCAGGGCGTGTGGCGCGTCTCCCAAGATCTCGGGCGTATCGCCGCGCAGAGCGG 870
Db 796 CAGCAGCGCGACCGGGCGCTCGCAAGATCTCGGGCGTATCGCCGCGCAGAGCGGCG 855
QY 871 GCATCGCCACGCCATGAAGTGGCCAGAGCACAGTGGAGCCTCGCAAGAACACGCTTT 930
Db 856 CACGAGCGCGGTACACGATGGCTCCGCCAGGCTGTTTCGAGCTCGACCCGACGCGCATG 915
QY 931 GAGCTCTACGGGCTGACTTCTCTTGGGAGGACTTCAGGCCCTTGGCTGATCGAGATC 990
Db 916 GCGGCGCGCTGGCGACGTCATGCGCGGAAGTGACCATGCGGGGCGAGCTCATGTGCG 975
QY 991 AATTCAGCCCCACCATGCACCCGCTCCACGCGCGTCCAGCGCGGAGCTGTGTGCACAGGTG 1050
Db 976 GACGCGCGCGACCGCGAGCGCGGACAGCTGTTC---GCCCGGTTCTCCGCGTGGCG 1032
QY 1051 CAGGAGGACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTC 1110
Db 1033 GAGCGCGCGCGGTGTACACGCGAGGAGACTACGCGCAACTCGTCGAGCACTTCGTGCGG 1092
QY 1111 CTGTGAGGCGACCGGTGTGAGCGCGCCCATTCAGCGGCTCCGACCTCTGCGTGGCG 1170
Db 1093 AGTGTGCGGCTGTGCGAGCTCGCGCGGGGCTCTCCGGGAGGGCGCGCGCAGGAG 1152
QY 1171 GCGCTCAGTGTGAGGAGACCGAGGAGGAGGTGCTG 1206
Db 1153 TACCTGTGCGGTTGCGGCCCAAGATCCGAGGATG 1188

RESULT 8
US-09-902-540-983
; Sequence 983, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 983
; LENGTH: 9039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-983

Query Match      3.1%; Score 50.2; DB 4; Length 9039;
Best Local Similarity 46.9%; Pred. No. 0.062;
Matches 190; Conservative 0; Mismatches 213; Indels 2; Gaps 1;

QY 1089 TGACATCGCAACTTCGAGCTCCTGTGGAGGCGAGCGGTGTGAGCGCCGCCATTCAG 1148
Db 2988 TGACCTCGCTCTTCGATATCCCGAGGGGTGCCAAGGAGCGTCCCGCGGTGAGCAA 3047
```


QY 861 CATGAAGAGGCCATGCGCCACGCCCATGAAGTGGCCCGGACGACCAAGTGGAGCCTCGCAA 920
Db 1617 CTTGGGCTACGCGCTTCGCCGAGTTCAGGAGCAGCAGCGCCTGAAGGCCCTCGCGCT 1676
QY 921 GAACAGC 927
Db 1677 GATCAAC 1683

RESULT 11
US-09-098-487-5
; Sequence 5, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-098-487-5

Query Match 3.0%; Score 48.6; DB 2; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.093;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 681 CCTGGACAGCTGGACAGCGCCATCCACTGTGTGCAACAGCGGTCCAGAAGTACCTGAA 740
Db 1437 CCAGAAGCATCTTGTGAGCGCAGCCCGCTGTGCTGTGCAACACTGCCCAAGGCCGTGGA 1496
QY 741 GAATGATGTGGCGCGCAGCCCTGTGCGCCGACACATGTGGACCGACCAAGTT 800
Db 1497 CGACAAGCAGCTGCGCAAGTGTCTGTAGCGCCACCGGGGAGAGGGGTGCGCAT 1556
QY 801 CCAGGAGTACCTGCAGCGCCAGGGCGGTGCGCGCGTGTGGGGCAGGTCTATCTACCCGTC 860
Db 1557 CAAGGAGTCCGCGTGTATGCGGACCTGAAGGGCGGTGCACGGCAACATGAAGGGCCAGAG 1616
QY 861 CATGAAGAGGCCATGCGCCACGCCCATGAAGTGGCCCGGACGACCAAGTGGAGCCTCGCAA 920
Db 1617 CTTGGGCTACGCGCTTCGCCGAGTTCAGGAGCAGCAGCGCCTGAAGGCCCTCGCGCT 1676
QY 921 GAACAGC 927
Db 1677 GATCAAC 1683

RESULT 12
US-08-471-033-18
; Sequence 18, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
US-08-471-033-18

Query Match 3.0%; Score 48.2; DB 1; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.11;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 507 CCACCCCTCTTTCCAGGACAAAGTGGGTGCCAGAGTACATCGAGACGCCGTGCT 566
Db 720 CCACCCCATCACCTCGAACAAGAGCGGTGGACACCTGCTGAACAAGCCCATGAT 779

QY 567 CATCTGTGACCAAGTTTCAGCATCAGACAGTGGTTCCTCGTCAGCGACTGGAAACCCCT 626
Db 780 GCTGGAGACCAACAGACCGGCGTCTCAAGATCAAGACACCCACCGCAACATCGT 839
QY 627 GACCATCTGGTTCTCAAGGAGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGA 686
Db 840 GACCGCGCGAGTGAACCGCGTGATCCAGCAGATCAAGGCCAAGACCCGCGAGCATCAT 899
QY 687 CAAGCTGGACAGCGCCATCACCCTGTGTGCAACAAGCGCGTCAGAAAGTACCTGAAGATGA 746
Db 900 CGTCGACGACGGCGAGCGCGTGGCGCGAGAGCGGTGGCGCCCAAGGACTACGAGAAACCC 959
QY 747 TGTGGCGCGAGCCCTGCTGTCGCGCGCACACAATGTGGACAGCAGCAGGTTTCCAGGA 806
Db 960 CGAGGACAAGACCCCGAGCTGACCTGAGAGAGCGCCCTGAAGCTGAGCTACCCCGACGA 1019
QY 807 GTACCTGCAGCGCCAGCGCGCGTGGCGCGTGTGGGGCAGCGTCTATCTACCCGTCATGAA 866
Db 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAGAACAGCCCATCTACGAGAGCAGCGT 1079
QY 867 GAAGGCATCGCCACCGCATGAAGTGGCGCCAGGACACAGTGGAGCGCTCGCAAGAACAG 926
Db 1080 GATGACCTATCTAGACGAGAACACCGCCAAAGAGGTGACCAAGCAGCTGAACGACACAC 1139
QY 927 CTTTGACCTACGGGCTGACTTCTGCTTGGAGGAGCTTCAAGGCGCTGAGTATCGA 986
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
QY 987 GATCA 991
Db 1200 GACCA 1204

RESULT 13

US-08-471-044-18
; Sequence 18, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868 Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,044
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018

; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA: US 08/037,057
; APPLICATION NUMBER: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SQLv3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
; US-08-471-044-18

Query Match 3.0%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.11;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 507 CCACCTCTTTCCAGGAGCAACAAGTGGTGGTCCAGAAATACATCGAGACGCGCTGCT 566
Db 720 CCACCCCATCACCTTGAACAAGAGCGGTGGCAACCTCTGAACAACAGCCCATGAT 779
QY 567 CATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAGCGACTGGAAACCCCT 626
Db 780 GCTGGAGACCAACAGACCGCGCTCTACAAGATCAAGGACACCCACGCAACATCGT 839
QY 627 GACCATCTGGTTCTCAAGGAGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGA 686
Db 840 GACCGCGCGAGTGAACCGCGTGATCCAGCAGATCAAGGCCAAGACCCGCGAGCATCAT 899
QY 687 CAAGCTGGACAGCGCCATCCACCTGTGCAACAAGCGCGTCCAGAAAGTACCTGAAGATGA 746
Db 900 CGTCGACGACGGCGAGCGCGTGGCGCGAGAGCGCGCCCAAGGACTACGAGAAACCC 959
QY 747 TGTGGCGCGAGCCCTGCTGCGCGCACACAATGTGGACAGCAGCAGGTTTCCAGGA 806
Db 960 CGAGGACAAGACCCCGAGCTGACCTGAAAGAGCGCCCTGAAGCTGAGCTACCCCGACGA 1019
QY 807 GTACCTGCAGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGTCTATCTACCCGTCATGAA 866
Db 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAGAACAGCCCATCTACGAGAGCAGCGT 1079
QY 867 GAAGGCATCGCCACCGCATGAAGTGGCGCCAGGACCAAGTGGAGCGCTCGCAAGAACAG 926
Db 1080 GATGACCTATCTAGACGAGAACACCCGCAAGGAGGTGACCAAGCAGCTGAACGACAC 1139
QY 927 CTTTGAGCTTACGGGCTGACTTCTGCTTGGAGGAGCTTCAAGGCGCTGAGTATCGA 986
Db 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGAGCTGAAGCTGACCCCAAGATGAACGT 1199
QY 987 GATCA 991
Db 1200 GACCA 1204

RESULT 14

US-08-463-483A-18
; Sequence 18, Application US/08463483A
; Patent No. 5849870
; GENERAL INFORMATION:

APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-463-483A-18
Query Match 3.0%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.11;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Db 840 GACCGCGCGGAGTGGAACGGCGTGATCCAGCAGATCAAGGCCAAGACCGCCAGCATCAT 899
QY 687 CAAGCTGGACAGCGCCATCCACCTGTGCAACAAAGCGCGTCCAGAAGTACTCTGAAGAATGA 746
Db 900 CGTCGACGACGGCGAGCGCGTGGCGAGAAAGCGGTGGCGCCGCAAGGACTTACGAGAACCC 959
QY 747 TGTGGCGCGCAGCCCCCTCTGCTGCCCGCACACAACATGTGGACCAAGCAGGTTCCAGGA 806
Db 960 CGAGGACAAAGACCCCCAGCCTGACCCCTGAAGGAGCGCCCTGAAGCTGAGCTACCCCGACGA 1019
QY 807 GTACCTGGACGCCAGGGCGGTGGCGCGCTGTGGGCGACGGTCTATCTACCGCTCCATGA 866
Db 1020 GATCAAGGAGATCGAGGGCGCTCTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCGT 1079
QY 867 GAAGGCCATCGCCACGCGCATGAAGGTGGCCCGCAGGACCAACGCTGGAGCCTTCGAAGAACAG 926
Db 1080 GATGACCTATCTAGACGAGAACACCGCCAAAGAGGTGACCAAGCAGCTGAACGACACCAC 1139
QY 927 CTTTGAGCTCTACGGGGCTGACTTCGTCTCTTGGGAGGGACTTCAGGCCCTTGGCTGATCGA 986
Db 1140 CGGCAAGTTCAAGGAGCTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
QY 987 GATCA 991
Db 1200 GACCA 1204
RESULT 15
US-08-471-046A-18
Sequence 18, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
TITLE OF INVENTION: Protein Genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5866326artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:


```
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLv4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPA(a) 80 kd protein from AB78"
US-08-471-046A-18

Query Match      3.0%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.11;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Qy 507 CCACCTCTTTCCAGGAGCAACAAGTGGGTGCTCCAGAAGTACATCGAGACGCGCGTGTCT 566
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 720 CCACCCCATCACCTGAAACAGAGAGAGTGGACAACTCTGTAACAAACAAGCCCATGAT 779
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 567 CATCTGTGACACCAAGTTTCACATCAGACAGTGGTCTCTGTCACGGACTGGAGACCCCT 626
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 780 GCTGGAGACCAACAGACCAGCGCGGTCTCAAGATCAAGGACACCCACCGGCAACATCGT 839
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 627 GACCATCTGGTTCTACAAGGAGAGTTACTTGGGTTTCTCAACTCAGCGCTTCTCCCTTGA 686
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 840 GACCGCGCGAGTGGAAACGCGGTGATCCAGCAGATCAAGGCCAAGACCCGACGATCAT 899
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 687 CAAGCTGGACAGCGCCATCCACCTGTGCAACAACGCCGTCCAGAAGTACCTGAAGAATGA 746
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 900 CGTCGACGACGCGGAGCGCGTGGCCGAGAGCGGTGGCCGCAAGGACTACGAGAACCC 959
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 747 TGTGGCCGCGAGCCCTCTGTCGCCGACACACATGTGGACCAACAGGTTCCAGGA 806
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 960 CGAGGACAAGACCCCTGAGCTGAGGAGCGCCCTGAAGCTGAGCTACCCCGACGA 1019
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 807 GTACCTGACGCGCAGGCGCGTGGCGCGTGTGGGCGAGCGTCTATCTACCCGTCATGAA 866
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCGT 1079
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 867 GAAGGCCATCGCCACGCCATGAAGTGGCCAGGACCAAGTGGAGCCTCGCAAGAACAG 926
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1080 GATGACCTATCTAGACGAGAACACCGCAAGGAGGTGACCAAGCAGCTGAACGACACCC 1139
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 927 CTTTGAGCTCTAGGGGCTGACTTCTGCTTGGGAGGACTTCAGGCCCTGGCTGATCGA 986
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1140 CGGCAAGTTCAAGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGT 1199
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 987 GATCA 991
Db      |||||
Qy 1200 GACCA 1204
Db      |||||
```

Search completed: September 24, 2005, 11:30:35
Job time : 297.571 secs

This Page Blank (uspio)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
C	1	1179	72.6	1259	9	AY415398 Homo sapi
	2	1284.6	69.5	1257	9	AY415399 Pan trogl
	3	804.4	49.6	2969	3	AK030151 Mus muscu
	4	798.2	49.2	2720	7	AK029745 Mus muscu
	5	768	47.3	770	7	CR745100 CR745100
	6	753.8	46.4	1265	5	AY415400 Mus muscu
	7	443.6	27.3	775	5	BU937372 AGENCOURT
	8	421.4	26.0	4184	3	BS468175 Homo sapi
	9	399.6	24.6	777	7	CK597795 AGENCOURT
	10	399	24.6	2520	3	AK080321 Mus muscu
C	11	378.4	23.3	1023	4	BM808637 AGENCOURT
	12	360	22.2	1022	4	BM808516 AGENCOURT
	13	321	19.8	693	4	BM717061 UI-E-B30-
	14	316	19.5	861	7	CK653583 AGENCOURT
	15	307.6	19.0	896	6	CD515431 AGENCOURT
	16	298.4	18.4	762	6	CD627148 56047432J
	17	284.2	17.5	503	7	CK604006 AGENCOURT
	18	258.8	15.9	583	7	CN349423 170004243
	19	258.2	15.9	680	6	CB854808 UI-CF-BC1
	20	252.8	15.6	514	7	CR746623 CR746623
C	21	245.4	15.1	320	1	AA144756 mr73d04.r
	22	222.4	13.7	429	7	W31827 ZB94704.r1
	23	219.2	13.5	880	5	BX448465 BX448465
	24	213.2	13.1	616	7	CN349424 170006000


```
QY 61 AGCAAGCCACGGGACCAAGAGGAGGAGGCGCGGAGCAGCAGCTTGAGCAGCAGCAAGAT 120
DB 78 AGCAAGCCACGGGACCAAGAGGAGGAGGCGCGGAGCAGCAGCTTGAGCAGCAGCAAGAT 137

QY 121 GCTGAAATCTGAGGCAAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 180
DB 138 GCTGAAATCTGAGGCAAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 197

QY 181 AAGGTGTGCAGAGCCTACCTGGGGCAGCTGAGGAGCAGCAGCAGCAGCAGCAGCAGAT 240
DB 198 AAGGTGTGCAGAGCCTACCTGGGGCAGCTGAGGAGCAGCAGCAGCAGCAGCAGCAGAT 257

QY 241 GCCGTGAGGACCTCAGTGGGCGGAGTGGAGGAGCCTGACCCAGCAGTACTACTCCCTC 300
DB 258 GCCGTGAGGACCTCAGTGGGCGGAGTGGAGGAGCCTGACCCAGCAGTACTACTCCCTC 317

QY 301 GTTCATGGGATGCTTTCATCTCAATTCAGAAATTAATTTTCGAGTGGCAGGCTCTG 360
DB 318 GTTCATGGGATGCTTTCATCTCAATTCAGAAATTAATTTTCGAGTGGCAGGCTCTG 377

QY 361 CTGAATAGAAATCAGCTCTGTGAACCTTCAGACGAGCATTGACGGGCTCGGAACATCTGG 420
DB 378 CTGAATAGAAATCAGCTCTGTGAACCTTCAGACGAGCATTGACGGGCTCGGAACATCTGG 437

QY 421 ATTATAAGCCCGCGGCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 480
DB 438 ATTATAAGCCCGCGGCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 497

QY 481 GAGATCTGAGCTGGCAGCTGAGACACCTCTCTTTTCAGGGAACAAGTGGGTGGTC 540
DB 498 GAGATCTGAGCTGGCAGCTGAGACACCTCTCTTTTCAGGGAACAAGTGGGTGGTC 557

QY 541 CAGAAGTACATCAGAGCGCGCTCATCTGTGACACCAAGTTCACATCAGACAGTGG 600
DB 558 CAGAAGTACATCAGAGCGCGCTCATCTGTGACACCAAGTTCACATCAGACAGTGG 617

QY 601 TTCTCTGTCACGAGCTGGAAACCCCTGACCATCTGGTTCTTACAAGGAGAGTTACTTGGCG 660
DB 618 TTCTCTGTCACGAGCTGGAAACCCCTGACCATCTGGTTCTTACAAGGAGAGTTACTTGGCG 677

QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGGCATCCACTGTGCAACAAC 720
DB 678 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGGCATCCACTGTGCAACAAC 737

QY 721 GCCTGCCAGAGTACCTGAAGAGATGATGTGGGCGGAGCCCTGCTGCGCCGACACAAC 780
DB 738 GCCTGCCAGAGTACCTGAAGAGATGATGTGGGCGGAGCCCTGCTGCGCCGACACAAC 797

QY 781 ATGTGGACCAAGCAGGTTCCAGGAGTACTCTGACGCGCCAGGCGCTGCGCGCTGTGG 840
DB 798 ATGTGGACCAAGCAGGTTCCAGGAGTACTCTGACGCGCCAGGCGCTGCGCGCTGTGG 857

QY 841 GGCAGGCTCATCTACCGCTCATGAAGAGGCGCATCGCCGACGCCATGAAGTGGGCCAG 900
DB 858 GGCAGGCTCATCTACCGCTCATGAAGAGGCGCATCGCCGACGCCATGAAGTGGGCCAG 917

QY 901 GACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTTACGGGCTCAGCTTCGTCCTTGG 960
DB 918 GACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTTACGGGCTCAGCTTCGTCCTTGG 977

QY 961 AGGGAATTCAGGCGCTCGGCTGATCGAGATCAATTCAGAGCCCAACCATGACCCCGTCCAG 1020
DB 978 AGGGAATTCAGGCGCTCGGCTGATCGAGATCAATTCAGAGCCCAACCATGACCCCGTCCAG 1037

QY 1021 CCGGTCAAGCCCGCAGCTGTGTGACAGGTGAGGAGGACACCATGAAGTGGCCGTGGAC 1080
DB 1038 CCGGTCAAGCCCGCAGCTGTGTGACAGGTGAGGAGGACACCATGAAGTGGCCGTGGAC 1097

QY 1081 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1140
DB 1098 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1157
```

```
QY 1141 CCATTGAGCGGTCCGACCTCTGCGTGGCGGCGTCAGT 1179
DB 1158 CCATTGAGCGGTCCGACCTCTGCGTGGCGGCGTCAGT 1196

RESULT 2
AY415399 1257 bp DNA linear GSS 12-DEC-2003
LOCUS Pan troglodytes HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY415399
VERSION AY415399.1 GI:39771358
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 1257)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1257)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1257
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1257
/locus_tag="HCM5544"
gene
ORIGIN
Query Match 69.5%; Score 1128.6; DB 9; Length 1257;
Best Local Similarity 96.2%; Pred. No. 1.6e-238;
Matches 1134; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1 ATGGCATCCAGCATCCTCAAGTGGGTGGTCAAGCTGAGGAGCAGGAGCAGGAGCAGGAGCAG 60
DB 18 ATGGCATCCAGCATCCTCAAGTGGGTGGTCAAGCTGAGGAGCAGGAGCAGGAGCAGGAGCAG 77

QY 61 ACAGAGCCAGGAGCAGAGGAGGAGGAGGCGGGAGGAGGAGCAGGAGCAGGAGCAGGAGCAGAT 120
DB 78 AGCAAGCCAGGAGCAGAGGAGGAGGAGGCGGGAGGAGCAGCCTGAGCAGCAGGAGCAGAT 137

QY 121 GCTGAAATCTGAGGCAAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 180
DB 138 GCTGAAATCTGAGGCAAGAGCTCAGGGGCTCCCGGGGAGGCTTGTGGACATCGCGTGC 197

QY 181 AAGGTGTGCAGGCGCTTACCTGGGGCAGCTGAGGAGCAGCAGCAGCAGCAGCAGCAGAT 240
DB 198 AAGGTGTGCAGGCGCTTACCTGGGGCAGCTGAGGAGCAGCAGCAGCAGCAGCAGCAGAT 257

QY 241 GCCGTGAGGACCTCAGTGGGCGGAGTGGAGGAGCCTGACCCAGCAGTACTACTCCCTC 300
DB 258 GCCGTGAGGACCTCAGTGGGCGGAGTGGAGGAGCCTGACCCAGCAGTACTACTCCCTC 317

QY 301 GTTCATGGGATGCTTTCATCTCAATTCAGAAATTAATTTTCGAGTGGCAGGCTCTG 360
DB 318 GTTCATGGGATGCTTTCATCTCAATTCAGAAATTAATTTTCGAGTGGCAGGCTCTG 377

QY 361 CTGAATAGAAATCAGCTCTGTGAACCTTCAGACGAGCATTGACGGGCTCGGAACATCTGG 420
DB 378 CTGAATAGAAATCAGCTCTGTGAACCTTCAGACGAGCATTGACGGGCTCGGAACATCTGG 437

QY 421 ATTATAAGCCCGCGGCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 480
DB 438 ATTATAAGCCCGCGGCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGGAG 497

QY 481 GAGATCTGAGCTGGCAGCTGAGACACCTCTCTTTTCAGGGAACAAGTGGGTGGTC 540
DB 498 GAGATCTGAGCTGGCAGCTGAGACACCTCTCTTTTCAGGGAACAAGTGGGTGGTC 557

QY 541 CAGAAGTACATCAGAGCGCGCTCATCTGTGACACCAAGTTCACATCAGACAGTGG 600
DB 558 CAGAAGTACATCAGAGCGCGCTCATCTGTGACACCAAGTTCACATCAGACAGTGG 617

QY 601 TTCTCTGTCACGAGCTGGAAACCCCTGACCATCTGGTTCTTACAAGGAGAGTTACTTGGCG 660
DB 618 TTCTCTGTCACGAGCTGGAAACCCCTGACCATCTGGTTCTTACAAGGAGAGTTACTTGGCG 677

QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGGCATCCACTGTGCAACAAC 720
DB 678 TTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGGCATCCACTGTGCAACAAC 737

QY 721 GCCTGCCAGAGTACCTGAAGAGATGATGTGGGCGGAGCCCTGCTGCGCCGACACAAC 780
DB 738 GCCTGCCAGAGTACCTGAAGAGATGATGTGGGCGGAGCCCTGCTGCGCCGACACAAC 797

QY 781 ATGTGGACCAAGCAGGTTCCAGGAGTACTCTGACGCGCCAGGCGCTGCGCGCTGTGG 840
DB 798 ATGTGGACCAAGCAGGTTCCAGGAGTACTCTGACGCGCCAGGCGCTGCGCGCTGTGG 857

QY 841 GGCAGGCTCATCTACCGCTCATGAAGAGGCGCATCGCCGACGCCATGAAGTGGGCCAG 900
DB 858 GGCAGGCTCATCTACCGCTCATGAAGAGGCGCATCGCCGACGCCATGAAGTGGGCCAG 917

QY 901 GACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTTACGGGCTCAGCTTCGTCCTTGG 960
DB 918 GACCAAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTTACGGGCTCAGCTTCGTCCTTGG 977

QY 961 AGGGAATTCAGGCGCTCGGCTGATCGAGATCAATTCAGAGCCCAACCATGACCCCGTCCAG 1020
DB 978 AGGGAATTCAGGCGCTCGGCTGATCGAGATCAATTCAGAGCCCAACCATGACCCCGTCCAG 1037

QY 1021 CCGGTCAAGCCCGCAGCTGTGTGACAGGTGAGGAGGACACCATGAAGTGGCCGTGGAC 1080
DB 1038 CCGGTCAAGCCCGCAGCTGTGTGACAGGTGAGGAGGACACCATGAAGTGGCCGTGGAC 1097

QY 1081 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1140
DB 1098 CGCAGCTGTGACATCGCAACTTCGAGCTCTCTGTGAGGAGCAGCGGTGTTGAGCGGCC 1157
```



```
Qy 361 CTGAATAGAATACAGTCTGTGAACCCCTCAGACGGACATTCAGCGGCTCCCGGAACATCTGG 420
Db 378 CTGAANNNTTCANNNTNNNNCTNAGACGACATTCAGCGNNCCGGAACATCTGG 437
Qy 421 ATTATAAGCCCGCGGCAAGTCCCGGCGCGGAGACATAGTGTGCATGGACCGTGTGGAG 480
Db 438 ATTATAAGCCCGCGGCAAGTCCCGGCGCGGAGACATAGTGTGCATGGACCGTGTGGAG 497
Qy 481 GAGATCTGAGCTGGAGCTGACAGACCCCTCTTTCAGAGGACAAACAAGTGGGTGGTC 540
Db 498 GAGATCTGAGCTGGAGCTGACAGACCCCTCTTTCAGAGGACAAACAAGTGGGTGGTC 557
Qy 541 CAGAAGTACATCGAGACGCGCTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
Db 558 CAGAAGTACATCGAGACGCGCTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 617
Qy 601 TTCCTCGTCCAGGACTGGAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTCGGG 660
Db 618 TTCCTCGTCCAGGACTGGAACCCCTGACCATCTGGTTCTACAGGAGAGTTACTTCGGG 677
Qy 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGACAGCGGCATCTGCTGTGCAACAAC 720
Db 678 TTCTCAACACAGCGCTTCTCCCTGGACAAAGCTGACAGCGGCATCTGCTGTGCAACAAC 737
Qy 721 GCCGTCCAGAGTACCTGGAAGTATGATGGCGCGCAGCCCTGCTGCGCGCACACAAC 780
Db 738 GCCGTCCAGAGTACCTGGAAGTATGATGGCGCGCAGCCCTGCTGCGCGCACACAAC 797
Qy 781 ATGTGGACACAGCAGGTTCCAGAGTACCTGAGCGCCAGGCGCGCTGGCGCGTGTGG 840
Db 798 ATGTGGACACAGCAGGTTCCAGAGTACCTGAGCGCCAGGCGCGCTGGCGCGTGTGG 857
Qy 841 GGCAGCGTCTATCTACCGTCCATGAAGAGCCATATCGCCACGCGCATGAAGGTGGCCAG 900
Db 858 GGCNNNGTCTATCTACCGTCCATGAAGAGCCATATCGCCACGCGCATGAAGGTGGCCAG 917
Qy 901 GACACGCTGAGCCTCCAGACAGCTTTGAGTCTACGGGGTGACTCGCTCTTGGG 960
Db 918 GACACGCTGAGCCTCCAGACAGCTTTGAGTCTACGGGGTGACTCGCTCTTGGG 977
Qy 961 AGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGAGCCCAACATGACCCGCTCCAG 1020
Db 978 AGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGAGCCCAACATGACCCGCTCCAG 1037
Qy 1021 CCGGTACGCGCCAGCTGTGTGACAGGTGCAGAGGACACCATCAAGGTGGCGGTGGAC 1080
Db 1038 CCGGTACGCGCCAGCTGTGTGACAGGTGCAGAGGACACCATCAAGGTGGCGGTGGAC 1097
Qy 1081 CGCAGCTGTGACATCGGCACTTCGAGTCTGTGGAGGACGCGGTGGTGGCCGCC 1140
Db 1098 CGCAGCTGTGACATCGGCACTTCGAGNCCCTGTGGAGGNNNNNGTGGTGGAGCGGCC 1157
Qy 1141 CCATTACGCGGTCGACCTCTGGTGGCGGCGTCAGT 1179
Db 1158 CCATTACGCGGTCGACCTCTGGTGGCGGCGTCAGT 1196

RESULT 3
AK030151
LOCUS AK030151 2969 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4933401B17 product:hypothetical tubulin-tyrosine
ligase containing protein, full insert sequence.
ACCESSION AK030151
VERSION AK030151.1 GI:26326134
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
```

```
JOURNAL MEDLINE 303, 19-44 (1999)
PUBMED 99279253
REFERENCE 10349636
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N.,
Okazaki, Y., Saito, K., Saito, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
FEATURES
Location/Qualifiers
1..2969
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="C57BL/6J"
/db_xref="FANTOM DB:4933401B17"
/db_xref="taxon.10090"
/clone="4933401B17"
/sex="male"
```


/tissue type="testis"
/clone lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"
490_-2835
/notes="unnamed protein product; hypothetical
tubulin-tyrosine ligase containing protein
(InterPro|IPR004344, evidence: InterPro)
putative"
/codon start=1
/protein_id="BAC26811.1"
/db_xref="GI:26326135"
/translation="MNGERKLUSSLTSDGDHKEENLKQIGIPDILSSSPKLDRYKIA
RQTEKAIKERKIFSIYGHYPVIRLRRKNGWEKFNFPKALQNLGSEDKSAETKE
NOIALERFDDIHVNSRLVKNIEPYLLMTIKEDVDYHSLRTCDQMLNHYGKTASFTT
KIGLCNMRSLPWYQANPNTFPFCKGLCTSEKOEFLDDPRRTVAASILKVVVLHQ
NYCSKVKGSKKEEAKNPSPKDPENPDLPKLSGOVVDTPACKVCVAGLQLEHE
DIDVSASTEALSEEWNDLTQOYILLVHGNASITDSKSYFAGCQALLSKISVNPQT
EIDGIRNIWIKPAASGRDIVCMRKNVENSILVAASDQTTKDNKWWQVYIETPML
IYDKFDIRQWFLVTDWNPITWFKESYLRFSRQSLDKLSDLSAHLNCSNITQRRLK
NDKERSPLPLPCHNMTSTFQEYLQKRGGTWSIYIPSMKRAVTNMRVAQDHVEA
RKNSPFLYADFTLGRDFKPLWTEINSSPTMHPSTPTVAQLCAQOVEDTIKVVDRLK
DRNCDLGNFELLWRQPAVELPPENGSDLCVEGIVSVKAKKQMPPIASVGLSLLDAP
PKVRSARALMETVIRPRTVRODWKREAKVLSLTWSPVMDAEVGRGAKPIYAEV
NIDQVNDKSHKSGITRVQSSKVPGLTSAQHPALFAQTMKPTQMTSSPPPTASGNH
RDSPPFCPIVFEELWLHPNSQRPRPSSCILQSRQAQWIRGIP"

Query Match 49.6%; Score 804.4; DB 3; Length 2969;
Best Local Similarity 74.8%; Pred. No. 7.5e-167;
Matches 1057; Conservative 0; Mismatches 336; Indels 21; Gaps 3;

QY 2 TGGCATCCAGCATCCTCAAGTGGGTGGTCAGCCACCAGAGCTGCGAGGAGCAGCAGAA 61
DB 1106 TGGCTGCCAGCATCCTCAAGTGGGTAGTCTCTCAACAGAAATTAC-----TGCAGCAAAAG 1159

QY 62 GCAAGCCAGGACACAGAGGAGAGGAGCCGGGAGCAGCGACTGAGCAGGAGCAGCAAGATG 121
DB 1160 TCAAGGGCAAGATAAGAGGAGGAGGCCAAGAACAGTGAACCCAGCCCCCAAGAAAGATC 1219

QY 122 CTGAAATCTGAGGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGGGACATCGCGTGCA 181
DB 1220 CTGAAATCTGAGCTCAAGCTACCAAGCTCTCGGGACAGGTTGTAGACACAGCTTGCA 1279

QY 182 AGGTGTGCCAGGCTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGATG 241
DB 1280 AGGTGTGCCAGGCTACCTGGGACAGCTAGAGCATGAAGACATAGACGTATCGGAGGCCA 1339

QY 242 CCGTGGAGGACCTACTAGGCGCAGTGGGAGGACCTGACCCAGCAGATGACTACTCCCTCG 301
DB 1340 GCACCGAGGCTTGTCTGAGGAGGAATGGAATGACCTGCACAGCAGTACTATCTACTGG 1399

QY 302 TTCATGGCGATGCTTTCATCTCAATTTCAAGAAATTTACTTTTCGAGTGCAGGCTCTGC 361
DB 1400 TTCATGGCAATGCTTTCATCTCAATCTGATTCGAAAGTTACTTTTGGCAATGCCAAGCTCTGC 1459

QY 362 TGAATAGATCACTGCTGTGAACCTCTCAGCGGACATTCAGCGGCTCCGGAAACATCTGGA 421
DB 1460 TGAGTAAGATCACTGCTGTGAACCTCTCAGACAGATCGATGGGATCCGGAACATCTGGA 1519

QY 422 TTATAAGCCCGGCGCAAGTCCCGGGGCGGAGACATAGTGTGCATGGAACCGTGTGGAGG 481
DB 1520 TCATAAAGCCTGCAGCAAGTCCCGGGTCGAGATATTTGTGTATGGACCGTGTGAGAGA 1579

QY 482 AGATCTGGAGCTGGCAGCTGCAGACACACCTCTTTCCAGGGACAACAAGTGGTGGTCC 541
DB 1580 ACATCTGAGTCTGTGGCGCGAGACAGCAGACACCAAGAGCAACAATGGGTGGTCC 1639

QY 542 AGAAGTACATCGAGAGCGCGCTGCTCATCTGTGACACCAAGTTTCAGACATCAGACAGTGT 601
DB 1640 AGAAGTACATCGAGACCCCAATGCTCATCTACGACACCAAGTTTTCAGACATCAGACAGTGT 1699

QY 602 TCCTCTGACGGAATGGAACCCCTGACCATCTGGTTCTTACAGGAGAGTACTTCGCGT 661

ORIGIN

1700 TCTCTGTCTCAGACTGGAATCCCTTAACCATCTGTTTCTACAAAGAAAGACTACTCTCGCT 1759
QY
662 TCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGTCAACAACG 721
DB
1760 TCTCCACACAGCGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGTAACAACT 1819
QY
722 CCGTCCAGAAAGTACTTGAAGAATGATGTGGGCGGACGCCCCCTGTGTGCCCGCACAAACA 781
DB 1820 CCATCCAGAGGGCGCTCAAAAATGATAAAGAGCGAGTCCGCTGTCTACCTTGTCTATAACA 1879

QY 782 TGTGGACACGACACAGGTTTCAGGAGTACTCTCAGCGCCGAGCGGCGGTGGCGCGTGTGG 841
DB 1880 TGTGGACACGACACTCGTTCCTCAGGAGTACTCTCAGGAGGCGGCGGAGAGGAACGTGG 1939

QY 842 GCAGCGTCACTTACCGCTCCATGAAGAAGCCATCGCCACGCCCATGAAGGTGGGCCAGG 901
DB 1940 GTAGCATCATCTACCGCTCTATGAAAAGAGCTGTCAACATGCCATGAGGTAGCCAGG 1999

QY 902 ACCAGTGGAGCCTCGCAAGAAACAGCTTTGAGCTCTA CGGGGCTGACTTTCGTCTTGGGA 961
DB 2000 ACCAGTGAAGCCCGTAAGAACAGCTTCGAGCTCTACGGAGCCGACTTCATCTCTGGGC 2059

QY 962 GGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCGACCATGCAACCGCTCCACGC 1021
DB 2060 GAGACTTCAAGCCATGGCTTATCGAGATCAATCCAGCGCCCATGCAACCCCTCCACTC 2119

QY 1022 CGGTACCGGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGGCGG----- 1075
DB 2120 CGGTACAGCTCAGCTCTGTGCCAGGTGCAGGAGGACACCATCAAGGTGGTGGATC 2179

QY 1076 -----TGAACCGCAGCTGTGACATCGGCAATTTCGAGCTCTGTGGAGGAGCGCGTGG 1129
DB 2180 GCAAACTGGACCGCAAACTGTGACATCGGCAACTTTGAGCTTCTGTGGCGGACGCTGCCG 2239

QY 1130 TTGAGCGCGCCCATTCAGCGGTCGAGCTCGACCTCTGCGTGGCGGCGCTCAGTGTGAGAG 1189
DB 2240 TGGAGTGCACACCGTTTCAACCGGCTGACCTCTGTGTGAAAGGTATCAGCGTGAAGAAAG 2299

QY 1190 CCAGGAGGAGGCTGTGTCGCCGTCTGCAACTCAAGCGCTCGGCTCGCTGTTGGAGCGCGC 1249
DB 2300 CCAAGAGCAATGCGGCCCATTTGCGAGGTGCGGCTCTCAGAACTACTTTAGATGC-- 2357

QY 1250 AGCGCTGAAGGCAAGGCGCCCTCGGCCATGCCAGACCTGCCAGGAGACCCCATCAC 1309
DB 2358 -TCCACCCAAAGTGGCGAGTGC CGGGCCCTGATGAAACAGTGTAGCGCTCCAGAA 2416

QY 1310 CAGCTCTCCAGCGGACTTGGGACTGAGAGGAGAGAGGGGCTCCCTCGGCTGCTCG 1369
DB 2417 CGACCGTGGCGAGGACTGGAACCGGAGAGCAAGGTACTTTCAACCACTTGTGTCCA 2476

QY 1370 CACCCCTTAAGGGGGCGAGCGCGGAGCGGTGGAGC 1403
DB 2477 TGCAGTAATGATGAGAGGTCAGAGGTGCGAGC 2510

RESULT 4
AK029745 2720 bp mRNA linear HTC 03-APR-2004
LOCUS
DEFINITION
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:490524K07 product:hypothetical Tubulin-tyrosine
ligase containing protein, full insert sequence.
AK029745
ACCESSION
AK029745.1 GI:26325681
VERSION
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
Carninci, P. and Hayashizaki, Y.
AUTHORS
High-efficiency full-length cDNA cloning
TITLE
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253


```
Qy 722 CCTCCAGAGTACCTGAGAAATGATGTGGGCGCAGCCCCCTGTGCGCCGACACAA 781
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1823 CCATCCAGAGCGCCTCAAAAATGATAAAGCGCAGTCGCTGCTACCTTGTCAATA 1882
Qy 782 TGTGGACCAAGCAGGTTCCAGAGTACTCTGAGCGCCAGGCGCTGGCGCGTGTGG 841
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1883 TGTGGACCAAGCAGTCTGTTTCCAGAGTACTCTGAGAGAGGGCGCAGGAGAACGTGG 1942
Qy 842 GCAGCGTCTACTACCGTCCATGAAGAGCCATCGCCACGCGCATGAAGGTGGCCAG 901
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1943 GTAGCATCTACCGCTCTATGAAGAAGCTGTACCAATGCCATGAGGTAGCCAGG 2002
Qy 902 ACCAGTGGAGCTCCGCAAGAACAGCTTTGAGCTCTACGGGGTGTACTTGTCTTTGGA 961
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2003 ACCACGTAGAAGCCGTGAAGACAGCTTCAGCTCTACGAGCGGAGCTTCATCTGGGG 2062
Qy 962 GGGACTTCAGGCGCTCGTGATCGAGATCAATTCAGAGCCCAACCATGCCCGTCCAGC 1021
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2063 GAGACTTCAAGCCATGGCTTATCGAGATCAACTCCAGAGCCCAACCATGCCACCTCCACTC 2122
Qy 1022 CGGTACGGGCCAGCTGTGTGCAAGGTGCGAGGAGCACCAATCAAGGTGGCG----- 1075
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2123 CGGTACAGCTAGCTCTGTGCCAGGTGCGAGGAGCACCAATCAAGGTGGTGGATC 2182
Qy 1076 -----TGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGCGACCGGTGG 1129
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2183 GCAAACTGGACCGAAACTGTGACATCGGCAACTTTGAGCTTCTGTGGCGGAGCCTGCCG 2242
Qy 1130 TTGAGCGCGCCCATTTTCAGGGGTCCGACTCTGCTGGGGCGGCTCAGTGTGAGAGAG 1189
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2243 TGGAGCTGCCACCGTTTCAACGGGTCTGACCTCTGTGTGGAAGGTATTCAGCGTGAAGAA 2302
Qy 1190 CCAGGAGGAGGTGTCGCCGTCTGCAACTCAAGGCTCGGCTCGCTGTTGGAGCGCG 1249
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2303 CCAGAGCAATATGCCGCCCATTTGCCAGGTGGCTCTCAGATCATTAGATGC-- 2360
Qy 1250 AGCCGTGAAGGCACGGGGCCCTCGGCCATGCCAGACCTGCCAGGAGACCCCATCAC 1309
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2361 -TCCACCCAAAGTGCAGGTGCCGGGCCCTGATGAAACAGTGATCAGGCTCCCAAGAA 2419
Qy 1310 CAGCTCTCAGCGGAGCTTGGGCTGAAGAGAGAGAGGGGCTCCCGCTGGCTTGGTGG 1369
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2420 CGACCGTGGCGGAGGACTGGAACCGGGAAGAACAAAGGGCTTCGCTGTGATCTCOAGC 2479
Qy 1370 CACCCCTAAAGGGGGGAGCGGAGAGCGGTGGAGCGCCAGACCCACCGCCCAAGAGCTG 1429
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2480 GCGTGCCTAGTTGTGTGGCTTCGCCGCCGCCCTTGGCCCTGCAAGCTCTCCCT 2539
Qy 1430 CTGGGAAGTGGAGTCCCGGCGCTGCCCTGTGCG 1464
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2540 GCGCCTTGATGCCGAGGGCGCCTTCTGTGCTGC 2574
```

```
RESULT 5
CR745100 770 bp mRNA linear EST 30-AUG-2004
LOCUS CR745100 Soares testis NHT Homo sapiens cDNA clone IMAGp97102270 ;
DEFINITION IMAGE:1645239 5', mRNA_sequence.
ACCESSION CR745100
VERSION CR745100.1 GI:51667573
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 770)
JOURNAL Ebert,L., Heil,O., Hennig,S., Korn,B., Neubert,P., Partsch,E.,
AUTHORS Peters,M., Radelof,U. and Schneider,D.
TITLE I.M.A.G.E. cDNA Clone Collection
JOURNAL Unpublished (2004)
COMMENT Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
```

Email: www.rzpd.de
RZPD; IMAGp97102270.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
ML3r, Primer sequence: TTTCACAGGAAACAGCTATGAC.

FEATURES
source

Location/Qualifiers
1..770
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp97102270 ; IMAGE:1645239"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaudo."

ORIGIN

Query Match 47.3%; Score 768; DB 7; Length 770;
Best Local Similarity 99.7%; Pred. No. 6.9e-159;
Matches 768; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 251 ACCTCACTGAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGGCG 310
Db 1 ACCTCACTGAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGGCG 60
Qy 311 ATGCTTTTCATCTCCAATTCAAGAAATTAATTTTCGAGTGCAGGCTCTGCTGAATAGAA 370
Db 61 ATGCTTTTCATCTCCAATTCAAGAAATTAATTTTCGAGTGCAGGCTCTGCTGAATAGAA 120
Qy 371 TCACGTCTGTGAACCTTCAGACGGACATTCGAGCGGCTCCGGAACATCTGGATTATAAAGC 430
Db 121 TCACGTCTGTGAACCTTCAGACGGACATTCGAGCGGCTCCGGAACATCTGGATTATAAAGC 180
Qy 431 CCGCGGCCAAAGTCCCGGGGCCGAGACATAGTGTGATGGACCGTGTGGAGGAGATCCTGG 490
Db 181 CCGCGGCCAAAGTCCCGGGGCCGAGACATAGTGTGATGGACCGTGTGGAGGAGATCCTGG 240
Qy 491 AGCTGGCAGCTCAGACCAACCTCTTTTCAGGGAACAAGTGGGTGGTGCAGAGATACA 550
Db 241 AGCTGGCAGCTCAGACCAACCTCTTTTCAGGGAACAAGTGGGTGGTGCAGAGATACA 300
Qy 551 TCGAGACGGCGCTGCTCATCTGTGACACCAAGTTGACATCAGACAGTGGTTCTCTCGTCA 610
Db 301 TCGAGACGGCGCTGCTCATCTGTGACACCAAGTTGACATCAGACAGTGGTTCTCTCGTCA 360
Qy 611 CGGACTTGAACCCCTCGTACCATCTGGTCTTACAAGGAGAGTTACTTTGCGGTTCTCAACTC 670
Db 361 CGGACTTGAACCCCTCGTACCATCTGGTCTTACAAGGAGAGTTACTTTGCGGTTCTCAACTC 420
Qy 671 AGCGCTTCTCTCTGACCAAGCTGGAGCAGCGCATTCACCTGTGCAACAACCGCGTCCAGA 730
Db 421 AGCGCTTCTCTCTGACCAAGCTGGAGCAGCGCATTCACCTGTGCAACAACCGCGTCCAGA 480
Qy 731 AGTACTTGAAGATGATGTGGCGCGCAGCCCTCTGCGCCGACACACATGTGGACCA 790
Db 481 AGTACTTGAAGATGATGTGGCGCGCAGCCCTCTGCGCCGACACACATGTGGACCA 540


```

Qy 791 GCACCGAGTCCAGAGTACTCTGACGCGCCAGGCGCGTGGCGCGTGTGGGCGAGCGTCA 850
Db 541 GCACCGAGTCCAGAGTACTCTGACGCGCCAGGCGCGTGGCGCGTGTGGGCGAGCGTCA 600
Qy 851 TCTACCGGTCCATGAAGAAGGCCATCGCCACGCGCATGAAGTGGCCCGCCAGGACCACTGG 910
Db 601 TCTACCGGTCCATGAAGAAGGCCATCGCCACGCGCATGAAGTGGCCCGCCAGGACCACTGG 660
Qy 911 AGCCTCGCAAGAACAGCTTTGAGCTCTACCGGGGTGACTTGGTCTTTGGGAGGAGCTTCA 970
Db 661 AGCCTCGCAAGAACAGCTTTGAGCTCTACCGGGGTGACTTGGTCTTTGGGAGGAGCTTCA 720
Qy 971 GGGCTCGGTGATCGAGATCAATTCAGCCCGCCACGATGACCCCGTCCACG 1020
Db 721 NGCCCTGGGTGATCGAGATCAATTCAGCCCGCCACGATGACCCCGTCCACG 770

RESULT 6
LOCUS AY415400 1265 bp DNA linear GSS 12-DEC-2003
DEFINITION Mus musculus HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY415400
VERSION AY415400.1 GI:39771359
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1265)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferred nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
JOURNAL PUBMED
REFERENCE 2 (bases 1 to 1265)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment
FEATURES
source location/Qualifiers
1..1265
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1265
/locus_tag="HCM5544"

gene
ORIGIN
Query Match 46.4%; Score 753.8; DB 9; Length 1265;
Best Local Similarity 78.6%; Pred. No. 1e-155;
Matches 934; Conservative 0; Mismatches 237; Indels 18; Gaps 2;

Qy 2 TGGCATCCAGCATCCTCAAGTGGGTGTGACGCCAATCCAGAGCTGCGAGGAGCAGCAAA 61
Db 19 TGGCTGCCAGCATCCTCAAGTGGGTGTGACGCCAATCCAGAGCTGCGAGGAGCAGCAAA 72
Qy 62 GCAAGCCAGGACCAAGGAGGAGGAGCGCGGAGCAGGACCTTGAGCAGGAGCAGCATG 121
Db 73 TCAAGGCGCAAGAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 132
Qy 122 CTGAAATGCTGAGGCAAGCTCAGGCGCTCCCGGGGCGAGCTTGTGGACATCGCGTGCA 181
Db 133 CTGAAATGCTGAGGCAAGCTCAGGCGCTCCCGGGGCGAGCTTGTGGACATCGCGTGCA 192

```

```

Qy 182 AGTGTGCCAGGCGCTACTCTGGGCGAGCTGGAGGCAATGAGGACATCGACACGTCAGCAGATG 241
Db 193 AGTGTGCCAGGCGCTACTCTGGGCGAGCTGGAGGCAATGAGGACATCGACACGTCATCGAGGCCA 252
Qy 242 CCGTGGAGGACCTCAGTCAGGCGCGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCG 301
Db 253 GCACCGAGGCGCTTGTCTGAGGAGGAGTGGAAATGACCTGACACAGCAGTACTTACTACTGG 312
Qy 302 TTCAATGCGATGCTTTTCATCTCCAAATTAAGAAATTAATTTTCAGCAGTCCGAGGCTTCG 361
Db 313 TTCAATGCGATGCTTTTCATCTCCAAATTAAGAAATTAATTTTCAGCAGTCCGAGGCTTCG 372
Qy 362 TGAATAGATCAAGTCTGTGAAACCTCAGACGGGACATTTGACGGGCTCGGGAACATCTTGG 421
Db 373 TGAGTAGATCAGTTCTGTGAAACCTCAGACGGGACATTTGACGGGCTCGGGAACATCTTGG 432
Qy 422 TTATAAAGCCGCGCGCCCAAGTCCCGGGCGAGACATAGTGTGATGATGATGATGATGATGATG 481
Db 433 TCATAAAGCCTGAGCCCAAGTCCCGGGTGAATATTTGTGATGATGATGATGATGATGATGATG 492
Qy 482 AGATCTCTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACCAAGTGGGTGGTTC 541
Db 493 ACATCTCTGAGTCTGTGGCGCGAGACAGCCAGACCAAAAGGACCAAAATGGGTGGTTC 552
Qy 542 AGAAGTACATCGAGACGCGCGCTGCTATCTGTGACACCAAGTTCGACATCAGACAGTGGT 601
Db 553 AGAAGTACATCGAGACGCGCGCTGCTATCTGTGACACCAAGTTCGACATCAGACAGTGGT 612
Qy 602 TCCTCGTCAAGGAGTGGAAACCCCTGACCATCTGGTTCTACAGGAGAGTACTTGGCGT 661
Db 613 TCCTCGTCAAGGAGTGGAAACCCCTGACCATCTGGTTCTACAGGAGAGTACTTGGCGT 672
Qy 662 TCTCAACTCAGCGCTTCTCCCTCGAAGCTGGACAGCGCATCCACTGTGCAACAACG 721
Db 673 TCTCCACACAGCGCTTCTCCCTCGAAGCTGGACAGCGCATCCACTGTGTAACAAC 732
Qy 722 CGTCCAGAAAGTACTGGAAGATGATGTGGCGCGCAGCGCCCTGTCTGCCCGCACACAACA 781
Db 733 CCATCCAGAGGCGCTCAAAATATGATAAGAGCGCAGTCCGCTGTCTACCTTGTATAACA 792
Qy 782 TGTGGACCAAGCAGCGTTCAGAGGTACCTGAGCGCCAGCGCGCGCGCGCGCGCGCGCGTGG 841
Db 793 TGTGGACCAAGCAGCGTTCAGAGGTACCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGTGG 852
Qy 842 GCAGCGTCACTTACCGGTCCATGAAGAGCCATCGCCACGCGCATGAAGTGGCGCCAGG 901
Db 853 GTAGCATCATCTACCGGTCTATGAAAAGAGCTGTCAACATGCGATGAGGAGTGGCGCGCG 912
Qy 902 ACCAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGTGACTTCTGCTCTCTGGGA 961
Db 913 ACCAGTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGTGACTTCTGCTCTCTGGGG 972
Qy 962 GGGACTTTCAGGCGCTGCGTGAATCAATTCAGCGCCCAAGCATGCAAGCGCGTCCAGCG 1021
Db 973 GAGACTTTCAGCGCTGCGTGAATCAATTCAGCGCCCAAGCATGCAAGCGCGTCCAGCG 1032
Qy 1022 CGGTACCGCGCCAGCTGTGTGACAGGTGAGGAGGACACCATCAAGTGGCGCGCGCGCG 1075
Db 1033 CGGTACAGCTCAGCTCTGTGCCAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1092
Qy 1076 -----TGAACCCAGCTGTGACATCGGCAATTCGAGCTCTCTGTGAGGAGCGCGGTGG 1129
Db 1093 GCAAACTGGACCGGAACTGTGACATCGGCAATTCGAGCTCTCTGTGAGGAGCGCGGTGG 1152
Qy 1130 TTGAGCGCGCGCGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1178
Db 1153 TGGAGCTGCCACCGTTCAAGGGGTCTGACCTCTGTGTGGAAAGGTATTCAG 1201

```

RESULT 7

BU937372

LOCUS

775 bp

mRNA

linear

EST 18-OCT-2002

DEFINITION AGENCOURT_10519180 NIH_MGC_169 Mus musculus cDNA clone
IMAGE:6705256 5', mRNA sequence.
ACCESSION BU937372
VERSION BU937372.1 GI:24126191
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 775)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLC3002 row: g column: 16
High quality sequence stop: 569.
Location/Qualifiers
1. .775
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6705256"
/lab_host="PH10B (Ti-phase-resistant)"
/clone_lib="NIH MGC 169"
/note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI
(ggccatgatgcc); Site 2: SfiI (ggccctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGTATACGAGAGTGGCCATTACGCCGG-3' and
5'-ATTATAGAGCCGAGCGCCGACATG-dt(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 27.3%; Score 443.6; DB 5; Length 775;
Best Local Similarity 74.5%; Pred. No. 3.1e-87;
Matches 576; Conservative 0; Mismatches 185; Indels 12; Gaps 1;
QY 586 GACATCAGACAGTGGTTCTCTGTCACGAGCTGGAACCCCTGACCATCTGGTTCTACAAG 645
DB 1 GGCATCAGACAGTGGTTCTCTGTCACAGACTGGAATCCCTTAACCATCTGGTTCTACAAA 60
QY 646 GAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGAGTGGACGCGGCATC 705
DB 61 GAAAGCTACTCTCGCTTCTCCACACAGCGCTTCTCCCTGGACAGAGTGGACGCGGCATC 120
QY 706 CACCTGTGCAACAAACCCGTCAGAGAGTACCTCAAGAAATGATGTGGCGCCGACGCCCTCG 765
DB 121 CACCTGTGTAACTCCATCCATCAGAGCGCGCTCAAAATGATTAAGAGCGCATCGCTG 180
QY 766 CTGCCCGCACACAACATGTGGACACAGACACAGGTTCCAGAGTACTGTGAGCGCCAGGGC 825
DB 181 CTACCTTGTATAAATGTGGACACAGACTCGCTTTCCAGGAGTACCTTCGAGAAGAGGGGC 240
QY 826 CGTGGCGCGTGTGGGGCAGCGTCATCTACCCGTCCATCAAGAGGCCATCGCCACGGC 885
DB 241 CGAGGAGGAACGTGGGGTAGCATCATCTACCCGTCTATGAAAGAGCTGTGCACCAATGCC 300
QY 886 ATGAAGTGGCCCGAGGACCAACGTGGAGCGCTCGCAAGAACAGCTTTTGTAGCTCTACGGGGCT 945
DB 301 ATGAGGGTAGCCAGGACCAAGCTAGAGCCCGTAGAAGCAAGCTTTCAGCTCTACGAGGC 360
QY 946 GACTTCGTCTTGGGAGGGAGCTTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCAACC 1005

DB 361 GACTTCATCTGNGCGAGACTTCAAGCCATGGCTTATCGAGATCAACTCCAGCCCCACC 420
QY 1006 ATGCACCCGTCACAGCGCGGTTCAGGGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATC 1065
DB 421 ATGCACCCCTCACTCCCGTCACAGCTCAGCTGTGTGCCAGGTGCAGGAGGACACCATC 480
QY 1066 AAGGTGGCG-----TGACCCCAAGCTGTGAATCGGCAACTTCGAGTCTCTG 1113
DB 481 AAGGTGGTGGTGGATCGCAAACTGGACCCGAACTGTGACATCGGCAACTTTGAGCTTCTG 540
QY 1114 TGGAGCAGCCGTTGAGTTCAGCGCCCGCCATTCAGCGGGTCCGACCTCTCGTGGCGGGC 1173
DB 541 TGGCGGACGCTCGCGTGGAGCTGCCACCGTTCAACGGGTCTGACCTCTGTGTGGAAGGT 600
QY 1174 GTCAGTGTGAGGAGCCAGGAGGAGGTGCTGCCCGTCTGCAACCTCAAGGCTCTGGCC 1233
DB 601 ATCAGCGTGAAGAAGCCCAAGAAAGAAATGCCGCCATTTGCCAGCGTTCGGCTCTCAGAA 660
QY 1234 TCGCTGTTTGGAGCGCAGCGCTGGAAGGACCGGGGCCCTCGGCCATGCCAGACCTGCC 1293
DB 661 TCACCTCTTAGATGCTCCACCCAAAGTGGCGAGTGGCCCGGCCCTGTGAAACAGTGA 720
QY 1294 CAGGAGACCCCATCACCAGCTCTCCAGCGGAGCTTGGGACTGAGGAAGAGAA 1346
DB 721 TCAGGCTTCAAGAACGACACCGTGGCGGAGACTGGGAAACGGGGGAAGAAGCA 773
RESULT 8
HSM808322
LOCUS Homo sapiens mRNA; cDNA DKF2p686D076 (from clone DKF2p686D076).
DEFINITION BX648175
ACCESSION BX648175.1 GI:34367334
VERSION
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4184)
AUTHORS Ansonge W., Krieger, S., Regiert, T., Rittmuller, C., Schwager, B.,
Newes, H.W., Weill, B., Amid, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.
CONSRMT The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKF2p686D076) is available at the RZPD Deutsches
Resourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
<http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKF2p686D076>
Further information about the clone and the sequencing project is
available at <http://mips.gsf.de/projects/cdna/>.
FEATURES
Location/Qualifiers
1. .4184
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKF2p686D076Q"
/db_xref="taxon:9606"
/clone="DKF2p686D076"
/tissue_type="cervix"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
/note="hypothetical protein"
1. .4184
/gene="DKF2p686D076"
2529. .3587
gene
CDS

Query Match	26.0%;	Score 421.4;	DB 3;	Length 4184;
Best Local Similarity	63.8%;	Pred. No. 3.1e-82;		
Matches 677;	Conservative 0;	Mismatches 371;	Indels 13;	Gaps 2;
Qy	80	GGGAGGAGGCGGGAGCAGCAGCAGCTGAGCAGCAGGCAAGATGCTGAAATGCTTGAGCGAA	139	
Db	2391	GGGATCTGTACAGGCCCTCCCTATGTGTCAGGAGACAAGCAGCGCCAAAGAACAGGAGANA	2450	
Qy	140	AGCTCAG-GGGCCCTCCCGGGCGACTTGTGTGACATCGCGTGCAGAGTGTGTCAGGCGCTAC	198	
Db	2451	AACCCAGTGTGTCTCCACAGATTGTGTGATGAAGCTCTGTGTGCTGGAGAGTAC	2510	
Qy	199	CTGGGCGAGCTGGAGCATGAGGACATCGACACAGTCGACAGATGCCGTGGAGGACCTCACT	258	
Db	2511	CTTAGCAACTTGGCCACATGGACATCGACAAGACCTGGAGGCCCGCTGTACTCACTCACC	2570	
Qy	259	GAGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTATGCGGATGCTTTTC	318	
Db	2571	CCCGAGGCGTGGTCCCTTCTCCAGCGCTACTACCAAGTGGTCTCCACGAAGGGGCGAGAA	2630	
Qy	319	ATCTCCAAATCAAGAAATTACTTTTTCGAGTGCACAGCTCTGCTGTAATAGAAATCACTCT	378	
Db	2631	CTCAGGCACCTGCACATCAGGTCACGCTGTGAGGACATCTCTGACAGAGCTGCAGGCC	2690	
Qy	379	GTGAACCTTCAGACGGCACTTTGACGGGCTCCGGAAACATCTGGATTATAAAGCCCGCGGCC	438	
Db	2691	GTGGTACCCACAGATAGACATGGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGGAGCC	2750	
Qy	439	AAGTCCCGGGCCGAGACATAGTGTGTGATGACCCGTTGTGAGGAGAGATCTCTGGAGCTGGCA	498	
Db	2751	AAGTCCCGCGGACGAGGCATCATGTGCATGGACCACCTCGGAGGAGATGCTCAAGCTGGT	2810	
Qy	499	GCTGCGACACCACTCTTTCCAGGGCAACAAGTGGGTGCTCCAGAAGTACATCGACAGC	558	
Db	2811	AACGGCAACCCCGTGGTGATGAAGGACGGCAAGTGGGTGTCAGAAAGTATATTGACGG	2870	
Qy	559	CCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGG	618	
Db	2871	CCCCCTCTCATCTTTGGCACCAAGTTTGACCTCAGACAGTGGTTCCTGGTTAACTGACTGG	2930	
Qy	619	AACCCCTGACCACTCTGGTTCTACAAGGAGATTACTTGCGGTTCTCAACTCAGCGCTTC	678	
Db	2931	AACCCACTTACCGTGTGTTCTACCGGACAGCTATATCCGCTTTTCCACGACGCCCTTC	2990	
Qy	679	TCCCTGCAAGCTGGACAGGCCATCCACTGTGTGCAACAACCGCGTCCAGAGTACTCTG	738	
Db	2991	TCCCTGAAGAACCTGGACAACACTCAGTGCACCTGTGTCAACAACCTCCATCCAGAAGCACCTG	3050	
Qy	739	AAGAAATGATGTGGGCCGACGCCCTCTGCCGACACAACAATGTCGACACGACACACGAG	798	
Db	3051	GAGAACTCATGCCATCGGCATCCACTGCTTCCGCGACACAACATGTGGTCTAGCCAGAGG	3110	
Qy	799	TTCCAGGAGTACCTTGACCGCCAGGGCCGTGGCGCCGTGTGGGCGACGCTCATCTACCCG	858	
Db	3111	TTCCAGGCCACCTGCAGGAGATGGGTGCCCAAAATGCTTGGTCCACCATCATCTGTCCT	3170	
Qy	859	TCCATGAAGAGGCCATTCGCCACAGCCCATGAAAGGTGGCCCGACGACCACTGTGGAGCTTCG	918	
Db	3171	GGCATGAAGAGATGCTGTGATCCACGCACTTCAGACCTCCAGAGACACCGTGCAGTGTGCG	3230	

Matches 513; Conservative 0; Mismatches 144; Indels 7; Gaps 2;	
QY	2 TGGCATCCAGCATCTCTCAAGTGGTGGTCTCAGCCAGCAGCTGCAGCAGGACGACAGAA 61
Db	
Db	43 TGGTCCAGCATCTCTCAAGTGGTGGTCTCAGCCAGCAGCTGCAGCAGGACGACAGAA 96
QY	62 GCAAGCCGAGGACACAGAGGGAGGAGCCGGGAGCAGCAGCTGCAGCAGGACGACGATG 121
Db	
Db	97 CCAAGGGCAAGAGTAAGAAGGAGGAGGCCCAAGAAATGGTGACCCAGCCCAAGAAAGATC 156
QY	122 CTGAAATCTGAGGCAAGACTCAGGGGCTCCGGGGGAGCTTGGGATCGGCTGGTGA 181
Db	
Db	157 CTGATAATCTGACCTCAAGTTTACCAAGCCTCTCAGGACAGGTTGAGACACAGATGCA 216
QY	182 AGGTGTGCCAGGCTTACCTTGGGGGAGCTGCAGCATCAGGACATCGACAGCTCAGCAGATG 241
Db	
Db	217 AGGTGTGCCAGGCTTACCTTGGGACAGCTGCAGCATCAGGACATCGACAGCTGCAGCGCA 276
QY	242 CCGTGGAGCACTCTAGAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCTCCCTCG 301
Db	
Db	277 GCGCCGAGGCTTGTCCGAGGAGGAATGGAAGCCTGACACAACTACTACTTGTCTGG 336
QY	302 TTCTAGGCGCATCTTCTCATCTCCATTCAGAAATTTACTTTTCGAGTGCCAGGCTCTGC 361
Db	
Db	337 TTCTAGGCAATGCTTCCATCAGGATTCGAGAGTTACTTTGGCAGTCCCAAGCTCTGC 396
QY	362 TGAATAGATACGCTCTGTGAACCTCTCAGACGACATTCACGGGCTCGGAACATCTGGA 421
Db	
Db	397 TGAGTAAGATCAGCTCTGTGAACCCCGACAGATTGACGGGATACGGAACATCTGGA 456
QY	422 TTATTAAGCCCGGCAAGTCCCGGGCCGAGACATAGTGTGCATGGAACCTGTGGAGG 481
Db	
Db	457 TCATTAAGCCTCGAGCAAGTCCCGGGCCGAGATATTGTGTGCATGGACCGGTGGAGA 516
QY	482 AGATCTGAGCTGGCAGCTCGAGACCACTCTTTCAGGAGCAACAGTGGTGGTCC 541
Db	
Db	517 ACATCTGATCTGTGGGCCACAGACGACGACCAACCAAGGACCAACAAATGGTGGTCC 576
QY	542 AGAAGTACATCAGACGCGCGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGTG 601
Db	
Db	577 AGAAGTACATCAGACGCGCGTCTCATCTGTGACACCAAGTTCGACATCAGACAGTGTG 636
QY	602 TCCTGCTCAGGACTGGAACCCCTGACCATCTGTTTCTCAAGGAGAGTTACTTCGCGGT 661
Db	
Db	637 TCCTGCTCAGACTGGAATCCCTT-AAATCTGTTCTTCAAGAGAGCTACCTCGGCTC 695
QY	662 TCTC 665
Db	
Db	696 TCAC 699
RESULT 10	
AK080321	
LOCUS	
DEFINITION	
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630053H17 product:HOTTL PROTEIN homolog [Homo sapiens], full insert sequence.	
ACCESSION	
AK080321	
VERSION	
AK080321.1 GI:26348476	
KEYWORDS	
HTC; CAP trapper.	
SOURCE	
Mus musculus (house mouse)	
ORGANISM	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
REFERENCE	
1 Carninci,P. and Hayashizaki,Y.	
AUTHORS	
TITLE	
High-efficiency full-length cDNA cloning	
JOURNAL	
Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE	
99279253	
PUBMED	
10349636	
REFERENCE	
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
Genome Res. 10 (10), 1617-1630 (2000)	
20499374	
11042159	
3	
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.	
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer	
Genome Res. 10 (11), 1757-1771 (2000)	
20530913	
11076861	
4	
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
Functional annotation of a full-length mouse cDNA collection	
Nature 409, 685-690 (2001)	
5	
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
Nature 420, 563-573 (2002)	
6 (bases 1 to 2520)	
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.	
Direct Submission	
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)	
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.	
Please visit our web site for further details.	
URL:http://genome.gsc.riken.jp/	
URL:http://fantom.gsc.riken.jp/	
Location/Qualifiers	
1. 2520	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="CS7BL/6J"	
/db_xref="FANTOM DB:A630053H17"	
/db_xref="taxon:10090"	
/clone="A630053H17"	
/tissue type="thymus"	
/clone_lib="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="3 days neonate"	

FEATURES
source


```
CDS
84..2198
/note="unnamed protein product; HOTTL PROTEIN homolog
[ Homo sapiens ] (SPR[Q9UI99], evidence: FASTA, 85.8%ID,
93.4%length, match=821)
putative"
/codon_start=1
/protein_id="BAC37878.1"
/db_xref="GI:26348477"
/translation="MDTSPPFYAAGQGEASPPSLPVSISPSLQSRWNRNTPYLIWT
TRRDVLDKRLSLKQDQINHYRAGSFTTKVGLCLNRLNLPWFEDADADFFPRCYRLG
AEDDKAFIEDFLWLTARNVKLIVLVEKSSQISIQAREEAPEDTQKQKELVIT
VSSDFVDEALSAQCEHLSTAHKDIDKPNPLSLPDDWSOFQORYQIVHGEASLR
YLEVQVQRCEDILOQNVVPLDMEGRDNIWIKPKAKSRGRGICWMLRDLMLKLV
DCNPLMKDGKVIQKVERPLIIGTKEDLROWFLVTDWNLTVWPRDYSYRFTSTQ
PFSKLNLSVHLNCSNIQHLEASCHRPMLPFDNWSQRFQHLQZEDVPADWSS
VIVPGKAAVHQALQVDSQNVCKRSFELYGFDFQFQWLLIINASPTMPTST
AVTPKACAGVQADLTWRDLRDSRCDTGAPELIYKQPAVEPQTVGRLVVGSTI
KEFVPVGRHRTGVSRLPHLLTQCGSEKSDGSPHRSASRKNRAESLEHTEKPEP
AAVASVSGKGGKAPFPPSLKSAWLPSPRVHRPQGRVLRLOHDLVSGSKALSTTCKA
LMTLPKAVLMSPPHDKLAPSLMKPKGVGFELECCCTTWRVLSGGISEGHRQAA
PRPSAPGKLSLSTFCSKTEI"
ORIGIN
Query Match 24.6%; Score 399; DB 3; Length 2520;
Best Local Similarity 61.8%; Pred. No. 2, 6e-77;
Matches 658; Conservative 0; Mismatches 395; Indels 12; Caps 1;
QY 139 AGCTCAGGGGCTCCCGGGGAGCTGTGGACATCGCGTGCAGAGGTGTGCCAGGCTAC 198
DB 552 AGCTAGTGACCGTGCTCAGACATTTGTCGATGAGGCTCTGATGTCATGCCAGGAGCAC 611
QY 199 CTGGGCGAGCTGAGCATGAGGACATGACACATGTCAGCAGATCCGCTGGAGGACCTCACT 258
DB 612 CTTAGCAGCATACCCACAAAGACATGTGCAAGGACCCCAACTCCCGCTCTCACTCAGC 671
QY 259 GAGCGCGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTCATGCGCATGCTTC 318
DB 672 CCGCATGATGTGTCACAGTTCCTTCAGCGCTACTACCAATAGTTTCATGAGGGGAGAG 731
QY 319 ATCTCCAAATTCAGAAATTAATTTTCCGAGTCCAGGCTCTGCTGAATAGAAATCAAGTCT 378
DB 732 CTCAGATACCTGAGAGTCCAGGCTGCAAGTCCAGGCTGTGAAGACATCTCAGCAGCTTCAGAAC 791
QY 379 GTGAACCTTCAGACGACATGAGGGGCTCCGGAACATCTGGATTATAAGCCCGCGGCC 438
DB 792 GTGGTACCCAGTTAGACATGAGGGGGGATCGGAACATCTGGATCTGGAAGCCCGGAGCC 851
QY 439 AAGTCCCGGGCGGACACATAGTGTGATGACCGCTGTGGAGGAGATCTCGGAGCTGGCA 498
DB 852 AAGTCCCGAGGCCGAGGATTAATGTGATGAACCGCTGGATGAGATGCTGAAGCTGGTG 911
QY 499 GCTGCAGACCCCTCTTTTCAGGGAACAAGTGGGTGGTCCAGAAATGATCATCGAGACG 558
DB 912 GACTGCAACCCCATGCTCATGAAGGATGGCAAGTGGATCGTGCAAGATACATTGAGCGG 971
QY 559 CGGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCAGGACTGG 618
DB 972 CCGCTCTTATCTTCGCGCACCAAGTGTGACCTGAGACAGTGGTTCCTGGTGAAGTACTGG 1031
QY 619 AACCCCTGACCATCTGTTCTACAAGGAGATTACTTGGGTTTCTCAACTCAGGCGCTTC 678
DB 1032 AACCCACTACCGTGTGGTGTCTACGAGACAGCTACATTCGTTCTCCACAGGCCCTTC 1091
QY 679 TCCTGGACAAGCTGGACGCGCATCCACTGTGCAACAAACCGCGCTCCAGAAAGTACCTTG 738
DB 1092 TCCTTAAGAACTGGCAACTCTGTGCACTCTGTGTAACAACCTCCATCCAGACACTTG 1151
QY 739 AGAATGATGTGGGCGGAGCCCTCTGCTGCCGCGACACAACATGTGGACCAAGCACAGG 798
DB 1152 GAGGCTCTCTGTCAACGGGACCCGATGCTGCCCCAGATAACATGTGGTCCAGCCAGAG 1211
QY 799 TTCCAGGATACCTGAGCGCCGAGGCGGTGGCGCGCTGTGGGCGGCTCATCTACCGG 858
```

```
Db 1212 TTTTCCGGCCCACTTTCAGGAGGTAGATGCCCAAGGCGCTGGTCCAGCGCTCATGTCGCCA 1271
QY 859 TCATGAAGAGGCCATCGCCACCGCATGAAGGTGGCCAGGACCAACGCTGGAGCCTCGC 918
Db 1272 GGCATGAAGGCTGCTGTGATCCATGCGCTCCAGACCTCCCAAGATAACGCTGACGTCGCG 1331
QY 919 AGAAGACAGCTTTTGGAGCTCTACGCGGCTGACTTCGTCCTTGGGAGGAGACTTCAGGCGCTGG 978
Db 1332 AGGCCAGCTTTTGGAGCTCTATGGGGGAGACTTTGTGTTGGGGAAGACTTCCAGGCCCTGG 1391
QY 979 CTGATCGAGATCAATTTCCAGCCCCCACCACCGCTCCACGCGGCTTCAGCGGCCAGCTG 1038
Db 1392 TTGATTGAATCAATGCCAGCCCCACCACCTTCCACGCGCTGTCAGTGCCTGCGGCTC 1451
QY 1039 TGTGCAAGGTGAGAGGAGACAC-----CATCAAGGTGGCGCTGGAGCGGAGC 1086
Db 1452 TGTGCGCGGTGTGAGGACAGATACCTTCGCTGTGTGTCATCGACCGGCGCTGGAGCGGAGC 1511
QY 1087 TGTGACATCGGCAATTTCCAGAGCTCTGTGAGGACGACGCGTGTGAGCGCGCCCAATTC 1146
Db 1512 TGTGATACGGAGACCTTTGAGCTCATCTATAGCAGCCTGCTGTGAGGTGCTCAGTAC 1571
QY 1147 AGCGGTCGCGACCTTCGCTGGCGGCGCTCAGTGTGAGGAGAGCC 1191
Db 1572 GTGGGTATCGGCTCTAGTGGAGGCTCTACCATCAAGAAGCCC 1616

RESULT 11
BM808637
LOCUS
DEFINITION
  BM808637 1023 bp mRNA linear EST 05-MAR-2002
  AGENCOURT_6617786 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734232
  5', mRNA sequence.
ACCESSION
  BM808637
VERSION
  BM808637.1 GI:19125460
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 1023)
  NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS
  National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
  Unpublished (1999)
JOURNAL
  Contact: Robert Strausberg, Ph.D.
  Email: cgabbs-remail.nih.gov
COMMENT
  Tissue Procurement: Invitrogen
  cDNA Library Preparation: Life Technologies, Inc.
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM12739 row: d column: 09
  High quality sequence start: 6
  High quality sequence stop: 664.
  Location/Qualifiers
    1..1023
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:5734232"
      /tissue_type="hippocampus"
      /lab_host="DH10B"
      /clone_lib="NIH_MGC_124"
      /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcorV
      (destroyed); Site 2: NotI; RNA source: male hippocampus,
      age 27. Library is oligo-dT primed and directionally
      cloned (EcorV site is destroyed upon cloning). Average
      insert size 1.4 kb, insert size range 0.9-4 kb. Library is
      normalized and enriched for full-length clones and was
      constructed by C. Gruber (Invitrogen). Research Genetics
      tracking code 012."
FEATURES
  source
ORIGIN
```


Query Match		23.3%;	Score 378.4;	DB 4;	Length 1023;
Best Local Similarity		69.5%;	Pred. No. 8.1e-73;		
Matches		514;	Conservative	0;	Mismatches 226; Indels 0; Gaps 0;
QY	345	GCAGTCCGAGCTCTGCTGAATAGATCATCAGTCTGTGAACCTCTCAGACGGACATGTGCGG	404		
DB	49	GGCGTGTGAGGACATCTCTGCAGCAGCTGCAGGCGGTGTGACCCAGATAGATGGAAGG	108		
QY	405	GCTCCGGAACATCTGATATATAAGCCCGCCGCAAGTCCCGGGCCGAGACATAGTGTG	464		
DB	109	GGATCGCAACATCTGGATCGTGAAGCAGAGCCAAAGTCCCGGACGAGGATCATGTG	168		
QY	465	CATGACCCGTGTGGAGGAGATCCTGAGCTGGCAGCTGCAGACCAACCCCTCTTTCCAGGA	524		
DB	169	CATGACCACTGGAGGAGATCTGAAGCTGTGAACGGCAACCCCGTGTGATGAAGA	228		
QY	525	CAACAAGTGGTGGTCCAGAAGTACATCGAGAGCGGCTGTCTATCTGTGACACCAAGTT	584		
DB	229	CGGCAAGTGGTGGTGCAGAAGTATATTGAGCGGCGCCCTCTCTCATCTTTGGCACCAAGTT	288		
QY	585	GCACATCAGACAGTGTCTCTCGTCACGAGCTGGAACCCCTGACCATCTGTTCTACAA	644		
DB	289	TGACCTCAGACAGTGTCTCTGTAACCTGACTGGAACCCACTTACCGTGTGTTCTACCG	348		
QY	645	GGAGAGTTACTTTCGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAAGGCCAT	704		
DB	349	GCACAGCTATATCCGCTTTTCACGCGAGCCCTTCTCCCTGGAAGAACCTTGGACAACCTCAGT	408		
QY	705	CCACTGTGCACAACCGCGTCCAGAAGTACCTGGAAGATGATGTGGCGGCGAGCCCT	764		
DB	409	GCACCTGTGCAACAACTCCATCCAGAAGCACCTGGAGAACTCATGCCATCGGCATCCACT	468		
QY	765	GCTGCCCGCACACAATGTGACAGCACAGCTTCCAGGAGTACCTGACGCGCAGGG	824		
DB	469	GCTTCGCGCAGAACATGTGTCTAGCCAGAGTTTCCAGGCCCTACCTGACGAGATGG	528		
QY	825	CGGTGGCGCGTGTGGGCGAGCTCATCTACCCGCTCCATGAAGAAGGCCATCGCCACGC	884		
DB	529	TGCCCAAAATGTTGTCACCATCATCTGTGCTGGCATGAAGGATGCTGTGATCCACGC	588		
QY	885	CATGAAGTGGCCAGACCACTGTGAGCGCTCGAAGAACACTTGTAGCTCTACGGGC	944		
DB	599	ACTTCAGACCTCCCAAGGACACCTGCAGCTGTCGGAAGCCAGCTTGTAGCTCTATGGCGC	648		
QY	945	TGACTTCGTCTGTGGAGGACTTCAGGCGCTGCTGATCGAGATCAATTCCAGCCCCAC	1004		
DB	649	TGACTTCGTTCGGGGAGACTTCAGCCCTGGCTGATTGAGATCAACGCCAGCCCCAC	708		
QY	1005	CATGACCCCGTCCAGCCCGTCCAGCCCGCTGTGTGACAGGTGCGAGGAGACACCAT	1064		
DB	709	GATGCAACCCCTCCACAGCAGTCACTGCGCGGCTCTGTGCTGCGGTGCAAGCTGACACCT	768		
QY	1065	CAAGTGGCGGTGGACCGCA	1084		
DB	769	GCGCGTGTGTCATTGACCGGA	788		

RESULT 12
BM808516
LOCUS
DEFINITION
AGENCY 6617783 NIH_MGC_124 Homo sapiens cdna clone IMAGE:5734160
5', mRNA sequence.
ACCESSION
BM808516
VERSION
BM808516.1 GI:19125339
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1022)
AUTHORS
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12739 row: a column: 09
High quality sequence start: 78
High quality sequence stop: 658.
Location/Qualifiers
1..1022
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5734160"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV (destroyed); Site_2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

ORIGIN

Query Match		22.2%;	Score 360;	DB 4;	Length 1022;
Best Local Similarity		68.8%;	Pred. No. 9.4e-69;		
Matches		495;	Conservative	0;	Mismatches 225; Indels 0; Gaps 0;
QY	345	GCAGTGCAGGCTGTGCTGAATAGAAATCACGTCTGTGAACCTCTCAGACGGACATGTGCGG	404		
DB	40	GGCGTGTGAGGACATCTCTGCAGCAGCTGCAGGCGCTGGTATCCAGATAGACATGAAGG	99		
QY	405	GCTCCGGAACATCGGATTATTAAGCCCGCGGCGCAAGTCCCGGGCCGAGACATAGTGTG	464		
DB	100	GGATCGCAACATCTGGAATCGTGAAGCCAGGAGCCAAAGTCCCGCGCAGGAGCATCATGTG	159		
QY	465	CATGACCCGTGTGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTTTCCAGGA	524		
DB	160	CATGACCACTGTGAGGAGATGCTGAAGCTGGTGAACCGCAACCCCGTGGTATGAAGA	219		
QY	525	CAACAAGTGGTGTGCAAGAAGTACATCGAGACGCGCTGTCTCATCTGTGACACCAAGTT	584		
DB	220	CGGCAAGTGGTGTGCGAGAAGTATATTGAGCGGCGCCCTCTCTCATCTTTGGCACCAAGTT	279		
QY	585	CGACATCAGACAGTGTTCCTCGTCACGAGCTGGGAACCCCTGACCATCTGGTTCACAA	644		
DB	280	TGACCTCTTACAGTGGTTCCTGGTAACTGACTGGGAACCCACTTACCGTGTGGTCTTACCG	339		
QY	645	GGAGAGTTACTTTCGCGTTCTCAACTCAGCGCTTCTCCTCGACAAGCTGGACAGGCCCAT	704		
DB	340	CGACAGCTATATCGCTTTTCCAGCGAGCCCTTCTCCTCGAAGAACCTGGACAACTCAGT	399		
QY	705	CCACTGTGCAACCAACGCGCTGCCAAGAAGTACCTGAAGAATGATGTGGGCGCGCAGCCCT	764		
DB	400	GCACCTGTGCAACCAACTCCATCCAGAACACCTCGGAGAACTCATGCCATCGGCATCCACT	459		
QY	765	GCTGCCCCGACACACATGTGGAACAGCACAGGTTCCAGGAGTACCTGCGAGCGCCAGG	824		
DB	460	GCTTCCGCGCAGACAACATGTGGTCTAGCCAGAGGTTCCAGGCCCCACCTGCGAGAGATGG	519		
QY	825	CCGTGGCGCGTGTGGGCGAGCGTCTATCTACCCGCTCCATGAAGAAGGCCATCGCCACGC	884		
DB	520	TGCCCAAAATGCTTGGTCCACCATCATCTGCTGCTGCAAGGATGCTGTGATCCACGC	579		
QY	885	CATGAAGTGGCCCGGACGACCGAGCTGCGCAAGAACAGCTTTGAGCTCTACGGGC	944		

Db 580 ACTTCAGACCTCCAGGACACCGTGCAGTGTGGAGGCCAGCTTTGAGCTCTATGGCG 639
 QY 945 TGACTTCGTCTTGGAGGAGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCCCAC 1004
 Db 640 TGACTTCGTCTTGGAGGAGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCCCAC 699
 QY 1005 CATGCACCCGCTCAGACCGGCTCAGCGCCAGCTGTGACAGGTGTCAGGAGGACACCAT 1064
 Db 700 GATGGCACCCCTCACAGCAGTCACTGGCCGGCTCTGTGCTGGCGTGAAGCTGACACCC 759

RESULT 13
 BM717061
 LOCUS
 DEFINITION
 UI-E-EJ0-ahk-e-05-0-UI-r2 UI-E-EJ0 Homo sapiens cDNA clone
 UI-E-EJ0-ahk-e-05-0-UI 5', mRNA sequence.
 BM717061
 VERSION
 BM717061.1 GI:19030319
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 693)
 AUTHORS
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL
 Genome Res. 6 (9), 791-806 (1996)
 MEDLINE
 97044477
 PUBMED
 8889548

COMMENT
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.
 Location/Qualifiers

FEATURES
 source
 1..693
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-ahk-e-05-0-UI"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DHI08 (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-EJ0"
 /notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-EJ0 is a subcloned cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT7T3-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes,
 AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment,
 AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina
 Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
 library was created for the program, Gene Discovery in the

Visual System, supported by National Eye Institute (NEI)."
 ORIGIN
 Query Match 19.8%; Score 321; DB 4; Length 693;
 Best Local Similarity 69.1%; Pred. No. 3.7e-60;
 Matches 473; Conservative 0; Mismatches 200; Indels 12; Gaps 2;
 QY 451 CGAGACATAGTGTGCATGACCGGTGTGGAGGAGATCCTGGAGTGGCAGCTGCAGACCAC 510
 Db 1 CGAGGACATCATGTGCATGACACCACTGGAGGAGATGCTGAAGCTGGT-GAACGGCAGCCC 59
 QY 511 CCTCTTTCCAGGACAAACAGTGGGTGGTCCAGAGTACATCGAGAGCCGCTGCTCATC 570
 Db 60 GTGGTCAATGAGACCGGCAAGTGGGTGGTGCAGAGTATATTGAGCGGCCCTCTCTCATC 119
 QY 571 TGTGACACCAAGTTCACATCAGACAGTGTCTCGTCACGAGTGGAAACCCCTTGACC 630
 Db 120 TTTGGCACCAGTTTGACCTCAGACAGTGTCTCGTAACTGACTGGAACCCACTTACC 179
 QY 631 ATCTGGTTCTACAGGAGAGTTCATTGGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAG 690
 Db 180 GTGTGGTTCTACCGGACAGCTATATCCGCTTTCCACGAGCCCTTCTCCCTGAAGAAC 239
 QY 691 CTGGACAGCGCCATCCACCTGTGCAACACGCGTCCAGAGTACCTGGAAGATGATGTG 750
 Db 240 CTGGACACTCAGTGCACCTGTGCAACAACTCATCCAGAGCACCCTGGAGAACTCATGC 299
 QY 751 GGCAGCAGCCCTCTGCTGCCCGCACACAATGTGGACAGCACCAGGTTTCCAGGAGTAC 810
 Db 300 CATCGGCATCCACTGCTTCCGCGCAGACAACATGTGTCTAGCCAGAGGTTTCCAGGCCAC 359
 QY 811 CTGCAGCGCCAGGCCGTGGCGCGTGTGGGAGAGTATCTACCGTCCATGAAGAAG 870
 Db 360 CTGCAGGAGATGGGTGCCCAAAATGTTGTGTCACCATCATCGTGTGCTGGCATGAAGGAT 419
 QY 871 GCCATCGCCAGCCCATGAAGTGGCGCCAGGACACAGTGGAGCTCGCAAGAACAGCTTT 930
 Db 420 GCTGTGATCCAGCACTTCAGACCTCCAGGACACCGTGCAGTGTGGAGGCGCAGCTTT 479
 QY 931 GAGCTCTACGGGCTGACTTCGTCTTGGAGGAGACTTCAGGCCCTGGCTGATCGAGATC 990
 Db 480 GAGCTCTATGGCGTGCATTCGTGTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATC 539
 QY 991 AATTCAGGCCACCATGCACCGCTGCACCGCGTCCAGCGCCAGCTGTGTGCACAGGTG 1050
 Db 540 AAGCCAGGCCCCAGCATGGCACCCCTCCACAGCAGTCACTGCCCGGCTGTGTGTGGCGTG 599
 QY 1051 CAGGAGGACAC-----CATCAAGTGGCCCTGGAGCGCAGCTGTGACATCGGCA 1099
 Db 600 CAAGCTGAACCTGCGCGTGGTGCATTGACCGGATGCTGGACCGCAACTGTGACACAGGAG 659
 QY 1100 ACTTCAGACTCTGTGGAGGCGAGCC 1124
 Db 660 CCTTTGAGCTCATCTATAAGCAGGC 684
 RESULT 14
 CK653583
 LOCUS
 DEFINITION
 AGENCOURT.17673480 NIH_MGC.237 Rattus norvegicus cDNA clone
 IMAGE:7113394 5', mRNA sequence.
 CK653583
 VERSION
 CK653583.1 GI:41389106
 KEYWORDS
 EST.
 SOURCE
 Rattus norvegicus (Norway rat)
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE
 1 (bases 1 to 861)
 AUTHORS
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Unpublished (1999)

COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs-remail.nih.gov Tissue Procurement: Howard Jacobs cDNA Library Preparation: Express Genomics cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM14985 row: e column: 08 High quality sequence stop: 304. Location/Qualifiers	Db	491	ATTATCTTCGATCTCTATGGAGCTGACTTTAATCTGGGCGTGACTTATAGTCTCTGTGT	550
		Qy	981	GATCGAGATCAATTCCAGCCCCACCATGCACCGCTCCAGCCGGTCACGGCC	1032
		Db	551	CTTCAAGATCACTTCTACTCTGCTTGGCACTCTACTTTGTAACCAACC	602
RESULT 15					
CD515431					
LOCUS	CD515431				
DEFINITION	AGENCOURT_14364765 NIH_MGC_181 Homo sapiens cDNA clone				
	IMAGE:30407505 5', mRNA sequence.				
ACCESSION	CD515431				
VERSION	CD515431.1				
KEYWORDS	EST.				
SOURCE	EST.				
ORGANISM	Homo sapiens (human)				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 896)				
AUTHORS	NIH-MGC http://mgi.nci.nih.gov/.				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. Michael Brownstein cDNA Library Preparation: Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: NDAM488 row: p column: 10 High quality sequence stop: 678. Location/Qualifiers				
FEATURES	1..896				
source	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:30407505" /tissue_type="White Matter" /dev_stage="Unknown" /lab_host="DH10B-Ton A (T1 and T5 phage resistances) " /clone_lib="NIH_MGC_181" /note="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning) . Average insert size 1.42 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."				
ORIGIN					
Query Match	19.0%; Score 307.6; DB 6; Length 896;				
Best Local Similarity	70.0%; Pred. No. 3.5e-57;				
Matches	434; Conservative 0; Mismatches 174; Indels 12; Gaps 1;				
Qy	520	AGGAGCAACAAGTGGTGTCCAGAGTACATCGAGACCGCGCTCATCTGTGACACC	579		
Db	25	AAAGGACGCGAGTGGGTGGTGAGAAGTATATGTAGCGCGCCCTCTCATCTTTGGCACC	84		
Qy	580	AAGTTTCGACATCAGACAGTGGTTCTCGTCACGAGTGGAAACCCCTGACCATCTCGTTTC	639		
Db	85	AAGTTTGACCTCAGACAGTGGTTCTGTGTAATGACTGGAACCCACTTACCGTGTGGTTC	144		
Qy	640	TACAGGAGAGTACTTTCGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGC	699		
Db	145	TACCGCGACAGCTATATCGCTTTTCCACGACGCTTCTCCCTGAAGAACCTGGACAAC	204		
Qy	700	GCCATCCACCTGTGCAACACGCGCTCCAGAGTACTTGAAGAATGATGTGGCGCCGACG	759		

Db 205 TCAGTGACCTGTGTCAACAACTCCATCCAGAACACCTGGAGAACTCATGCCATCGGCAT 264
Qy 760 CCCCTGCTGCCGACACACATGTGGACACAGACACAGGTTCAGGAGTACCTGCGAGCGC 819
Db 265 CCAGTCTTCCGCCAGACACATGTGGTCTAGCCAGAGGTTCAGGCGCCACCTGCGAGGAG 324
Qy 820 CAGGGCGGTGGCGCGTGTGGGCGAGCGTCTATCTACCCGTCCATGAAGAGGCCATCGCC 879
Db 325 ATGGTGCCCAAAATGCTTGGTCCACCATCTGTGCTGGCATGAAGGATGCTGTGATC 384
Qy 880 CACGCCATGAAGGTGGCCAGGACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTCTAC 939
Db 385 CACGCACTTCAGACCTCCAGGACACCGTGCATGTGGAAGGCCAGCTTTGAGCTCTAT 444
Qy 940 GGGGCTGACTTGTCTCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGC 999
Db 445 GGGGCTGACTTGTGTTCGGGGAGGACTTCAGGCCCTGGCTGATGAGATCAACGCCAGC 504
Qy 1000 CCCACCATGCACCCGTCCACGCCGTCCAGGCCAGCTGTGCACAGGTGCAGGAGGAC 1059
Db 505 CCCACGATGGCACCTCCACAGAGTCACTGCCCGGCTCTGTGCTGGCGTGCAAGCTGAC 564
Qy 1060 ACCATCAAGGTGGCC-----GTGGACCGCAGCTGTGACATCGGCAACTTCGAG 1107
Db 565 ACCCTGCGGTGTCTATTGACCGGAGGCTGGACCGCAACTGTGCACAGGAGCCTTGAG 624
Qy 1108 CTCCTGTGGAGGAGCGCGT 1127
Db 625 CTCATCTATAAGCAGCCCGT 644

Search completed: September 25, 2005, 01:56:01
Job time : 6239.23 secs

This page blank (uspto)

no.	score	match	length	db	id	description
1	1623	100.0	1838	12	ADJ93357	AdJ93357 Human BGS
2	1619.8	99.8	3554	12	ADJ93364	AdJ93364 Human tbs
3	1505.8	92.8	1859	12	ADJ93362	AdJ93362 Human BGS
4	1475.8	90.9	3465	12	ADJ93363	AdJ93363 Human BGS
5	1474.2	90.8	1939	12	ADJ93361	AdJ93361 Human BGS
6	726	44.7	726	12	ADJ93359	AdJ93359 Human BGS
7	501	30.9	101270	12	ADQ17814	Human sof
8	432.8	26.7	2380	6	AA598994	AdQ17814 Human sof
9	431.6	26.6	2326	6	AA598994	AA598994 Polynucle
10	427	26.3	490	12	AH16735	AH16735 Human cDN
11	421.4	26.0	2538	12	ACH91699	ACH91699 Human gen
12	421.4	26.0	2553	13	AD573168	AdS73168 Human kid
13	421.4	26.0	2881	3	ACN37881	ACn37881 Tumour-as
14	421.4	26.0	3001	4	AA158606	AAc77214 Human ORF
15	421.4	26.0	3001	5	ADQ98824	AdQ98824 Human pol
16	421.4	26.0	3001	9	ADB49584	AdB49584 DNA encod
17	419.8	25.9	3294	8	ABX34502	ABx34502 Novel hum
18	418	25.8	2848	11	ADM03081	AdM03081 Human cDN
19	413	25.4	2611	3	AA21812	AAf21812 Human bre
20	347.4	21.4	2412	12	ADH45468	AdH45468 Human mol

Federal TN Williams Nelson TC.

XX WPI; 2004-099381/10.
DR P-PSDB; ADJ93358.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
disorders.
XX
PS Claim 1; SEQ ID NO 1; 343pp; English.
XX
CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.
XX
SQ Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Query Match 100.0%; Score 1623; DB 12; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGGCCACGAGAGTGCAGAGGAGCAGCAGA 60
DB |||||
DB 153 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGGCCACGAGAGTGCAGAGGAGCAGCAGA 212
QY 61 AGCAAGCCAGGACCCAGAGGGAGGAGGCGGGAGCAGCGACCTGAGCAGCAGGCAAGAT 120
DB |||||
DB 213 AGCAAGCCAGGACCCAGAGGGAGGAGGCGGGAGCAGCGACCTGAGCAGCAGGCAAGAT 272
QY 121 GCTGAAATGCTCAGGCAGCAGCTCAGGGGCTCCCGGGCAGCTTGTGACATCGCGTGC 180
DB |||||
DB 273 GCTGAAATGCTCAGGCAGCAGCTCAGGGGCTCCCGGGCAGCTTGTGACATCGCGTGC 332
QY 181 AAGGTGTCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTTCAGCAGAT 240
DB |||||
DB 333 AAGGTGTCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTTCAGCAGAT 392
QY 241 GCCGTGGAGGACCTCACTAGGCGCAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 300
DB |||||
DB 393 GCCGTGGAGGACCTCACTAGGCGCAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC 452
QY 301 GTTCATGGCGATCTTTCATCTCCAATTCAGAAATTACTTTTCGAGTGCACAGGCTCTG 360
DB |||||
DB 453 GTTCATGGCGATCTTTCATCTCCAATTCAGAAATTACTTTTCGAGTGCACAGGCTCTG 512
QY 361 CTGAATAGAAATCAGCTCTGTGAACCTTCAGACGGACATTTGACGGGCTTCGGAAACATCTGG 420
DB |||||
DB 513 CTGAATAGAAATCAGCTCTGTGAACCTTCAGACGGACATTTGACGGGCTTCGGAAACATCTGG 572
QY 421 ATTATAAAGCCCGGCCAAGTCCCGGGGCCGAGACATAGTGTGCATGGACCCGTGTGGAG 480

DB 573 ATTATAAAGCCCGGCCAAGTCCCGGGGCCGAGACATAGTGTGCATGGACCCGTGTGGAG 632
QY |||||
QY 481 GAGATCCTCGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACAACAAGTGGGTGTC 540
DB |||||
DB 633 GAGATCCTCGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACAACAAGTGGGTGTC 692
QY |||||
QY 541 CAGAAGTACATCAGAGACGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 600
DB |||||
DB 693 CAGAAGTACATCAGAGACGCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGTGG 752
QY |||||
QY 601 TTCCTCGTCAAGAGTGGAAACCCCTGACCATCTGTTCTTACAAGGAGAGTTACTTTCGGG 660
DB |||||
DB 753 TTCCTCGTCAAGAGTGGAAACCCCTGACCATCTGTTCTTACAAGGAGAGTTACTTTCGGG 812
QY |||||
QY 661 TTCTCAACTCAGGGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 720
DB |||||
DB 813 TTCTCAACTCAGGGCTTCTCCCTGGACAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 872
QY |||||
QY 721 GCCGTCCAGAAGTACCTGAAGAAATGATGTGGGCGCGAGCCCTCTGTGCGCCGACACAAC 780
DB |||||
DB 873 GCCGTCCAGAAGTACCTGAAGAAATGATGTGGGCGCGAGCCCTCTGTGCGCCGACACAAC 932
QY |||||
QY 781 ATGTGACACGACACAGGTTCAGAGTACCTGACGCGCCAGGCGCGTGGCCGCTGTGG 840
DB |||||
DB 933 ATGTGACACGACACAGGTTCAGAGTACCTGACGCGCCAGGCGCGTGGCCGCTGTGG 992
QY |||||
QY 841 GCGAGGTCTATCTACCCGTCATGAAGAGGACATCGCCAGCCCATGAAGTGGGCCACAG 900
DB |||||
DB 993 GCGAGGTCTATCTACCCGTCATGAAGAGGACATCGCCAGCCCATGAAGTGGGCCACAG 1052
QY |||||
QY 901 GACACGTCGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTTGGG 960
DB |||||
DB 1053 GACACGTCGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTTGGG 1112
QY |||||
QY 961 AGGAGCTTCAGGCGCTGGCTGATTCAGATCAATTCAGGCCCCACATGACCCGCTCCAG 1020
DB |||||
DB 1113 AGGAGCTTCAGGCGCTGGCTGATTCAGATCAATTCAGGCCCCACATGACCCGCTCCAG 1172
QY |||||
QY 1021 CCGGTCACGGCCAGCTGTGCAGAGTGCAGAGGAGACCATCAAGTGGCCGCTGGAC 1080
DB |||||
DB 1173 CCGGTCACGGCCAGCTGTGTGCAGAGTGCAGAGGAGACCATCAAGTGGCCGCTGGAC 1232
QY |||||
QY 1081 CCGAGCTGTGACATCGGCAACTTCAGAGCTCTGTGTGAGGAGCAGCCGCTGTTGAGCGGCC 1140
DB |||||
DB 1233 CCGAGCTGTGACATCGGCAACTTCAGAGCTCTGTGTGAGGAGCAGCCGCTGTTGAGCGGCC 1292
QY |||||
QY 1141 CCATTCAGCGGGTCCGACCTCTGCGTGGCGGGCTCAGTGTGAGGAGAGCCAGGAGGACAG 1200
DB |||||
DB 1293 CCATTCAGCGGGTCCGACCTCTGCGTGGCGGGCTCAGTGTGAGGAGAGCCAGGAGGACAG 1352
QY |||||
QY 1201 GTGCTGCCGCTGTGCAACCTCAAGGCTCGGCTCGCTGTGAGAGCGCGAGCGCTGGAAG 1260
DB |||||
DB 1353 GTGCTGCCGCTGTGCAACCTCAAGGCTCGGCTCGCTGTGAGAGCGCGAGCGCTGGAAG 1412
QY |||||
QY 1261 GCACGGGGGCCCCCTCGGCCCATGCGAGACCTGCGCCAGGGACCCCCCATCACAGCTCTCCAG 1320
DB |||||
DB 1413 GCACGGGGGCCCCCTCGGCCCATGCGAGACCTGCGCCAGGGACCCCCCATCACAGCTCTCCAG 1472
QY |||||
QY 1321 CCGGACCTTGGGACTGAAGAGAGAGGGGCTCCCTCTGGCCCTTGTGTGCGACCTTTAAGG 1380
DB |||||
DB 1473 CCGGACCTTGGGACTGAAGAGAGAGGGGCTCCCTCTGGCCCTTGTGTGCGACCTTTAAGG 1532
QY |||||
QY 1381 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAGAGCTCTCGGAGAGTG 1440
DB |||||
DB 1533 GGGGAGCCGAGAGCGGTGGAGCCGACAGCCACCCGACCAAGAGCTCTGGGAAGGTG 1592
QY |||||
QY 1441 GAGCTCCCGGCTGCGCCCTGTGCGCACGCTGGAGAGTCAAGGCCCCCAACACACCGGTGTCCCC 1500
DB |||||
DB 1593 GAGCTCCCGGCTGCGCCCTGTGCGCACGCTGGAGAGTCAAGGCCCCCAACACACCGGTGTCCCC 1652
QY |||||
QY 1501 GTAGCCGAGCCCGCCAAAGCTGGGATCCAAACAGAGCTTAAATGGCGACCCGCTGGAGCT 1560
DB |||||

CC	tyrosine ligase modulators. In addition, the disclosed sequences may be
CC	useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC	used for diagnosing a pathological condition or a susceptibility to a
CC	pathological condition in a subject, and for preventing, treating or
CC	ameliorating a medical condition, such as a disorder related to aberrant
CC	tubulin ligase activity, a disorder related to aberrant tubulin-
CC	carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC	disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC	lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC	neural disorders, brain cancer, liver cancer, or proliferative condition
CC	of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC	polypeptide, polynucleotide, or their modulators are also useful for
CC	treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC	disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC	-42 polypeptide can be used as a preventive agent for immunological
CC	disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC	disease or scleroderma. The antibodies may be used to purify, detect and
CC	target the BGS-42 polypeptides. The present sequence is that of the
CC	tubulin tyrosine ligase protein consensus gene sequence which was used in
CC	the exemplification of the invention.
XX	
SQ	Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;
	Query Match 99.8%; Score 1619.8; DB 12; Length 3554;
	Best Local Similarity 99.9%; Pred. No. 0;
	Matches 1621; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGGCCACAGAGCTGCGAGCAGGACGACGAGA 60
DB	1695 ATGGCATCCAGCATCTCAAGTGGTGGTTCAGGCCACAGAGCTGCGAGCAGGACGACGAGA 1754
QY	61 AGCAAGCCACGAGGACACAGAGGGAGGAGCGCGGAGCAGCGACTGTAGCAGCAGCGCAAGAT 120
DB	1755 AGCAAGCCACGAGGACACAGAGGGAGGAGCGCGGAGCAGCGACTGTAGCAGCAGCGCAAGAT 1814
QY	121 GCTGAAAATGCTAGGCAAAAGCTCAGGGGCTCCCGGGCAGGCTTGTGACATCGCGTGC 180
DB	1815 GCTGAAAATGCTAGGCAAAAGCTCAGGGGCTCCCGGGCAGGCTTGTGACATCGCGTGC 1874
QY	181 AAGGTGTGCAGGSCCTACCTGGGCGAGCTGGAGCATGAGGACATCCACAGCTCAGCAGAT 240
DB	1875 AAGGTGTGCAGGSCCTACCTGGGCGAGCTGGAGCATGAGGACATCCACAGCTCAGCAGAT 1934
QY	241 GCGGTGGAGGACCTCACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 300
DB	1935 GCGGTGGAGGACCTCACTGAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 1994
QY	301 GTTCATGGCGATGCTTTTCATCTCCAATTCAGAAATTACTTTTCGACAGTCCAGGCTCTG 360
DB	1995 GTTCATGGCGATGCTTTTCATCTCCAATTCAGAAATTACTTTTCGACAGTCCAGGCTCTG 2054
QY	361 CTGAATAGAAATCAGCTCTGTGAACCTCTCAGCGGACATTTAGCGGGCTCCGGAACATCTGG 420
DB	2055 CTGAATAGAAATCAGCTCTGTGAACCTCTCAGCGGACATTTAGCGGGCTCCGGAACATCTGG 2114
QY	421 ATTATAAGCCCGCGCCCAAGTCCCGGGGCGGAGACATAGTGTGTCATGGACCGTGTGGAG 480
DB	2115 ATTATAAGCCCGCGCCCAAGTCCCGGGGCGGAGACATAGTGTGTCATGGACCGTGTGGAG 2174
QY	481 GAGATCTGTGAGCTGGCAGCTGCAGACCAACCTCTTTTCAGGGGACAAAGTGGGTGGTC 540
DB	2175 GAGATCTGTGAGCTGGCAGCTGCAGACCAACCTCTTTTCAGGGGACAAAGTGGGTGGTC 2234
QY	541 CAGAAGTACATCAGACCGCGCTCATCTGTGACACCAAGTTCGACATCCACAGTGG 600
DB	2235 CAGAAGTACATCAGACCGCGCTCATCTGTGACACCAAGTTCGACATCCACAGTGG 2294
QY	601 TTCCTCGTCACGGACTGGAAACCCCTCGACCATCTGGTCTTACAGGAGAGTACTATCGCG 660
DB	2295 TTCCTCGTCACGGACTGGAAACCCCTCGACCATCTGGTCTTACAGGAGAGTACTATCGCG 2354
QY	661 TTCCTCAACTCAGCGCTTCTCCCTGGGACAAAGCTGGACAGCGCCATCCACTGTGCACAAAC 720

Db 2355 TTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCCATCCACTGTGCAACAAC 2414
Qy 721 GCGTTCAGAAAGTACCTGAAGAATGATGTGGCGCGCAGCGCCCTGTGTCGCCGACACAAC 780
Db 2415 GCGTTCAGAAAGTACCTGAAGAATGATGTGGCGCGCAGCGCCCTGTGTCGCCGACACAAC 2474
Qy 781 ATGTGACCAAGCACCGAGTTCCAGAGTACTGTGAGCGCCAGCGCGTGGCGCGTGTGG 840
Db 2475 ATGTGACCAAGCACCGAGTTCCAGAGTACTGTGAGCGCCAGCGCGCGTGGCGCGTGTGG 2534
Qy 841 GGCAGCGTCTATCTACCGTTCATGAGAGAGGCGCATGCCCAACGCCATGAAGTGGGCCAG 900
Db 2535 GGCAGCGTCTATCTACCGTTCATGAGAGAGGCGCATGCCCAACGCCATGAAGTGGGCCAG 2594
Qy 901 GACACGTGAGCGCTCGCAAGAACAGCTTTGAGCTCTACGGGCGTCACTCGTCTTGGG 960
Db 2595 GACACGTGAGCGCTCGCAAGAACAGCTTTGAGCTCTACGGGCGTCACTCGTCTTGGG 2654
Qy 961 AGGACTTCAAGGCCCTGGCTGATCGAGATCAATTTCCAGGCCCAACCATGCAACCGTCCACG 1020
Db 2655 AGGACTTCAAGGCCCTGGCTGATCGAGATCAATTTCCAGGCCCAACCATGCAACCGTCCACG 2714
Qy 1021 CCGGTCAACGCCCTGCTGTGACAGGTGCAGAGGACACATCAAGTGGCGCTGGAC 1080
Db 2715 CCGGTCAACGCCCTGCTGTGACAGGTGCAGAGGACACATCAAGTGGCGCTGGAC 2774
Qy 1081 CGCAGCTGTGACATCGCACTTCGAGCTTCTGTGAGGCGCGCTGTGAGCGCGCTGG 1140
Db 2775 CGCAGCTGTGACATCGCACTTCGAGCTTCTGTGAGGCGCGCTGTGAGCGCGCTGG 2834
Qy 1141 CCATTCAAGCGGCTCGCACTTCGAGCTTCTGTGAGGCGCGCTGTGAGGAGCGAGCGAG 1200
Db 2835 CCATTCAAGCGGCTCGCACTTCGAGCTTCTGTGAGGCGCGCTGTGAGGAGCGAGCGAG 2894
Qy 1201 GTGCTCGCGCTGCAACCTCAAGGCTCGGCTCGCTGTGCAACGCGAGCGCTGAAG 1260
Db 2895 GTGCTCGCGCTGCAACCTCAAGGCTCGGCTCGCTGTGCAACGCGAGCGCTGAAG 2954
Qy 1261 GCACGGGGCCCTCGGCGCATGCCAGACCTTCCAGGGACCCCAACCAACAGCTCTCCAG 1320
Db 2955 GCACGGGGCCCTCGGCGCATGCCAGACCTTCCAGGGACCCCAACCAACAGCTCTCCAG 3014
Qy 1321 CGGACTTGGGACTGAAGGAAGAGAGGGCTCTCCCTGTGCTGGCACCCCTTAAGG 1380
Db 3015 CGGACTTGGGACTGAAGGAAGAGAGGGCTCTCCCTGTGCTGGCACCCCTTAAGG 3074
Qy 1381 GGGGACCGCAGAGCGTGGAGCGGCACAGCCCAACCGCACCAAGCTGCTGGGAAGTG 1440
Db 3075 GGGGACCGCAGAGCGTGGAGCGGCACAGCCCAACCGCACCAAGCTGCTGGGAAGTG 3134
Qy 1441 GAGCTCCCGGCTGCGCTGTGCGCACGTGGACAGTCAAGGCGCCCAACCAACCGGTCGCC 1500
Db 3135 GAGCTCCCGGCTGCGCTGTGCGCACGTGGACAGTCAAGGCGCCCAACCAACCGGTCGCC 3194
Qy 1501 GTAGCCGAGCCGCCAAAAGCTGGGATCCAAACCAAGCTAAATGCGCACCGCTGGAGCCT 1560
Db 3195 GTAGCCGAGCCGCCAAAAGCTGGGATCCAAACCAAGCTAAATGCGCACCGCTGGAGCCT 3254
Qy 1561 GTGCTCGGGGCTGAAGACAGAGAGGCGCGCTGCTCCGCGCCGAGGAAAAGT 1620
Db 3255 GTGCTCGGAGCCTGAAGACAGAGAGGCGCGCTGCTCCGCGCCGAGGAAAAGT 3314
Qy 1621 TCA 1623
Db 3315 TCA 3317

RESULT 3

ADJ93362

ID ADJ93362 standard; DNA; 1859 BP.

XX

AC ADJ93362;

XX

DT 06-MAY-2004 (first entry)
XX Human BGS-42 protein-related DNA clone B SeqID10.
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory; anti-inflammatory; anabolic; hypertensive;
KW neuroprotective; endocrine; antiparkinsonian; antiarthritic; antitubercular;
KW osteopathic; neurotropic; antiparkinsonian; antitubercular; antitubercular;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine kinase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary cancer; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
XX Homo sapiens.
XX WO2004005487-A2.
XX 15-JAN-2004.
XX 09-JUL-2003; 2003WO-US021605.
XX 09-JUL-2002; 2002US-0394725P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX Feder JN, Wu S, Nelson TC;
XX WPI; 2004-099381/10.
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
XX Example 4; SEQ ID NO 10; 343pp; English.
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antinflammatory, anabolic, hypertensive, osteopathic, neurotropic,
CC antiparkinsonian, antitubercular, antitubercular, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine kinase modulators. In addition, the disclosed sequences may be
CC used for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX SQ Sequence 1859 BP; 418 A; 563 C; 562 G; 316 T; 0 U; 0 Other;

Query Match 92.8%; Score 1505.8; DB 12; Length 1859;

Best Local Similarity 97.0%; Pred. No. 3.2e-311;

Matches 1574; Conservative 0; Mismatches 2; Indels 47; Gaps 2;			
Qy	1	ATGGCATCCAGCATCTCAAGTGGTGGTCAAGCCAGAGAGCTGCAGCAGAGCAGAGCAGA	60
Db	47	ATGGCATCCAGCATCTCAAGTGGTGGTCAAGCCAGAGAGCTGCAGCAGAGCAGCAGA	106
Qy	61	AGCAAGCCAGGAGCAGAGAGGAGGCGCGGAGCAGCAGCCTGAGCAGCAGCAGAGAT	120
Db	107	AGCAAGCCAGGAGCAGAGAGGAGGCGCGGAGCAGCAGCCTGAGCAGCAGCAGAGAT	166
Qy	121	GCTGAATAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC	180
Db	167	GCTGAATAATGCTGAGGCAAAAGCTCAGGGGCTCCCGGGGAGCTTGTGGACATCGCGTGC	226
Qy	181	AAGGTGTGCCAGGCTTACTTGGGACAGCTGAGAGCATGAGACATCGACATCGACAGAT	240
Db	227	AAGGTGTGCCAGGCTTACTTGGGACAGCTGAGAGCATGAGAGCATCGACATCGACAGAT	286
Qy	241	GCGTGGAGGACCTCACTGAGGGCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC	300
Db	287	GCGTGGAGGACCTCACTGAGGGCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTC	346
Qy	301	GTTTCATGGCGATGCTTTCATCTCCAATTCAAGAAATTAATTTTCGCAAGTGCAGGCTCTG	360
Db	347	GTTTCA-----TGGCAGGCTCTG 363	
Qy	361	CTGAATAGAAATCAAGTCTGTGAACCTCTCAGACGGAATTCAGCGGCTCCGGAAATCTGG	420
Db	364	CTGAATAGAAATCAAGTCTGTGAACCTCTCAGACGGAATTCAGCGGCTCCGGAAATCTGG	423
Qy	421	ATTATAAAGCCCGCGGCAAGTCCCGGGCGGAGACATAGTGTGCATGGACCGTGTGGAG	480
Db	424	ATTATAAAGCCCGCGGCAAGTCCCGGGCGGAGACATAGTGTGCATGGACCGTGTGGAG	483
Qy	481	GAGATCCTGGAGTGGCAGCTGCAGACCACTCTTTCCAGGGACAAAGTGGGTGTC	540
Db	484	GAGATCCTGGAGTGGCAGCTGCAGACCACTCTTTCCAGGGACAAAGTGGGTGTC	543
Qy	541	CAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG	600
Db	544	CAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGG	603
Qy	601	TTCTCTCAAGTGGAGTGGAAACCCCTGACCATCTGGTTTCAAGAGAGTACTTTCGCGG	660
Db	604	TTCTCTCAAGTGGAGTGGAAACCCCTGACCATCTGGTTTCAAGAGAGTACTTTCGCGG	663
Qy	661	TTCTCAACTCAGCGCTTCTCCCTGGACCAAGCTGGACAGCGCCATCCACCTGTGCAACAAC	720
Db	664	TTCTCAACTCAGCGCTTCTCCCTGGACCAAGCTGGACAGCGCCATCCACCTGTGCAACAAC	723
Qy	721	GCGGTCCAGAAATGATGTGGGCGGACAGCCCTGTGCTGCGCCGACACAAC	780
Db	724	GCGGTCCAGAAATGATGTGGGCGGACAGCCCTGTGCTGCGCCGACACAAC	783
Qy	781	ATGTGGACAGCAGGTTCCAGGATACCTGACGCGCCAGGCGCTGCGGCGGTGTGG	840
Db	784	ATGTGGACAGCAGGTTCCAGGATACCTGACGCGCCAGGCGCTGCGGCGGTGTGG	843
Qy	841	GGCAGCGCTCATCTACCGCTCATGAAGAAGCCATCCGCCACGCAATGAAGTGGCCCGAG	900
Db	844	GGCAGCGCTCATCTACCGCTCATGAAGAAGCCATCCGCCACGCAATGAAGTGGCCCGAG	903
Qy	901	GACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTTACCGGGGTGACTTCGCTTGGG	960
Db	904	GACCAAGTGGAGCTCGCAAGAACAGCTTTGAGCTTACCGGGGTGACTTCGCTTGGG	963
Qy	961	AGGCACTTCAGGCGCTGGCTGATCGAGATCAATTCAGCCGCCACCATGACCCGTCACG	1020
Db	964	AGGCACTTCAGGCGCTGGCTGATCGAGATCAATTCAGCCGCCACCATGACCCGTCACG	1023
Qy	1021	CCGGTCAAGCCCGAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGTGGCGGTGGAC	1080
Db	1024	CCGGTCAAGCCCGAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGTGGCGGTGGAC	1083

RESULT 4

ADJ93363

ID ADJ93363 standard; DNA; 3465 BP.

XX

AC ADJ93363;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human BGS-42 protein-related DNA clone C SeqID11.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nontropic; antiparkinsonian; antichratic; antiaethmatic;
anti-Hiv; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; ds.

Homo sapiens.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PF 09-JUL-2003; 2003WO-US021605.
XX
PR 09-JUL-2002; 2002US-0394725P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
PI WPI; 2004-099381/10.
XX
DR
XX
XX New testis-specific tubulin tyrosine-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Example 4; SEQ ID NO 11; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, anticiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Query Match 90.9%; Score 1475.8; DB 12; Length 3465;
Best Local Similarity 95.7%; Pred. No. 8.9e-305;
Matches 1554; Conservative 0; Mismatches 2; Indels 67; Gaps 1;

QY 1 ATGGCATCCAGCATCCTCAAGTGGGTGGTCCAGCCACGAGCTGCAGCAGGAGCAGCAGA 60
DB 1678 ATGGCATCCAGCATCCTCAAGTGGGTGGTCCAGCCACGAGCTGCAGCAGGAGCAGCAGA 1737

QY 61 AGCAAGCCCGAGGACACAGAGGAGGAGGCGCGGAGCAGCGACCTGAGCAGCAGGCGCAAGAT 120
DB 1738 AGCAAGCCCGAGGACACAGAGGAGGAGGCGCGGAGCAGCGACCTGAGCAGCAGGCGCAAGAT 1797

QY 121 GCTGAAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGCGCTTGTGGACATCGCGTGC 180
DB 1798 GCTGAAAATGCTGAGCAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGGACATCGCGTGC 1857

QY 181 AAGGTGTGTCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 240
DB 1858 AAGGTGTGTCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGAT 1917

QY 241 GCCGTGGAGGACCTACTAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 300
DB 1918 GCCGTGGAGGACCTACTAGGCGGAGTGGGAGGACCTGACCCAGCAGTACTACTCCCTC 1977

QY 301 GTTCATGGGAGTCTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGGCCAGGCTCTG 360
DB 1978 GTTCATGGGAGTCTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGGCCAGGCTCTG 2037

QY 361 CTGAATAGAAATCACGCTCTGTGAAACCTCTCAGACGGACATTTGACGGGCTCCGGAAACATCTCTG 420
DB 2038 CTGAATAGAAATCACGCTCTGTGAAACCTCTCAGACGGACATTTGACGGGCTCCGGAAACATCTCTG 2097

QY 421 ATTATTAAGCCCGCGCCAAAGTCCCGGGGCGGAGACATAGTGTGCATGACCGTGTGGAG 480
DB 2098 ATTATTAAGCCCGCGCCAAAGTCCCGGGGCGG----- 2129

QY 481 GAGATCCTGGAGCTGGCAGCTGCAGACCACTCTTTTCAGGAGCAACAAGTGGGTGGTC 540
DB 2130 -----AGGGACAAACAAGTGGGTGGTC 2150

QY 541 CAGAAGTACATCEGACGCGCTGTCTCATCTGTGACACCAAGTTCCACATTCAGACAGTGG 600
DB 2151 CAGAAGTACATCEGACGCGCTGTCTCATCTGTGACACCAAGTTCCACATTCAGACAGTGG 2210

QY 601 TTCTCTCGTCCAGCATCGGAAACCCCTTGACCATCTGGTTCTTACAAGGAGTACTTCTGCGG 660
DB 2211 TTCTCTCGTCCAGCATCGGAAACCCCTTGACCATCTGGTTCTTACAAGGAGTACTTCTGCGG 2270

QY 661 TTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 720
DB 2271 TTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGACAGCGCCATCCACCTGTGCAACAAC 2330

QY 721 GCCGTCCAGAAGTACTGAAGAATGATGTGGGCGCAGCCCCCTGTCTGCCCGCACACAAC 780
DB 2331 GCCGTCCAGAAGTACTGAAGAATGATGTGGGCGCAGCCCCCTGTCTGCCCGCACACAAC 2390

QY 781 ATGTGGACCAAGCAGCATCGTTTCAGGAGTACTTCAGCGCCAGCGGCGCTGGCCGCTGTGG 840
DB 2391 ATGTGGACCAAGCAGCATCGTTTCAGGAGTACTTCAGCGCCAGCGGCGCTGGCCGCTGTGG 2450

QY 841 GCAGCGCTCATCTACCCGCTCCATGAAGAGGCATCGCCACAGCCATGAAGTGGGCCCGCAG 900
DB 2451 GCAGCGCTCATCTACCCGCTCCATGAAGAGGCATCGCCACAGCCATGAAGTGGGCCCGCAG 2510

QY 901 GACCAAGTGGAGCCTCGCAAGAACAGCTTGTAGCTCTACGGGGCTGACTTGTGCTTGGG 960
DB 2511 GACCAAGTGGAGCCTCGCAAGAACAGCTTGTAGCTCTACGGGGCTGACTTGTGCTTGGG 2570

QY 961 AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCACCCATGACCGCTCCAG 1020
DB 2571 AGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCACCCATGACCGCTCCAG 2630

QY 1021 CCGGTCACGGCCCGAGCTGTGTCACAGGTGCAGAGGACACCATCAAGTGGCGCTGGAC 1080
DB 2631 CCGGTCACGGCCCGAGCTGTGTCACAGGTGCAGAGGACACCATCAAGTGGCGCTGGAC 2690

QY 1081 CGCAGCTGTGACATCGGCAACTTTCGAGCTCTGTGTGAGGAGCAGCCGGTGTGAGCGCGCC 1140
DB 2691 CGCAGCTGTGACATCGGCAACTTTCGAGCTCTGTGTGAGGAGCAGCCGGTGTGAGCGCGCC 2750

QY 1141 CCATTCAGCGGGTTCGACCTCTGCGTGGCGGGCGTCAGTGTGAGGAGACCCAGGAGGCGAG 1200
DB 2751 CCATTCAGCGGGTTCGACCTCTGCGTGGCGGGCGTCAGTGTGAGGAGACCCAGGAGGCGAG 2810

QY 1201 GTGCTGCCGCTGCAAACTCAAGGCTCGGCTCGCTGTGTGACCGCGCAGCGCTGAAG 1260
DB 2811 GTGCTGCCGCTGCAAACTCAAGGCTCGGCTCGCTGTGTGACCGCGCAGCGCTGAAG 2870

QY 1261 GCACGGGGCCCTCGGCGCATGCGACCTGTGCCAGGGACCCCATCAAGCTTCTCCAG 1320
DB 2871 GCACGGGGCCCTCGGCGCATGCGACCTGTGCCAGGGACCCCATCAAGCTTCTCCAG 2930

QY 1321 CGGGACTTGGGACTGAGGAAGAGGGGCTCCCGCTGGCTTGTGCTGCGACCCCTTAAGG 1380
DB 2931 CGGGACTTGGGACTGAGGAAGAGGGGCTCCCGCTGGCTTGTGCTGCGACCCCTTAAGG 2990

QY 1381 GGGGCGACCGCAGAGCGGTGGAGCGCGCACAGCCCAACCCGACCAAAAGCTGTCTGGGAAGGTG 1440
DB 2991 GGGGCGACCGCAGAGCGGTGGAGCGCGCACAGCCCAACCCGACCAAAAGCTGTCTGGGAAGGTG 3050


```
QY 721 GCGCTCCAGAGTACCTGAAGATGATGTGGCCGAGCGCCCTGTCGCCGCACACAAC 780
Db |||||
806 ACCGTCCAGAGTACCTGAAGATGATGTGGCCGAGCGCCCTGTCGCCGCACACAAC 865
QY 781 ATGTGACAGACACAGGTTCCAGGAGTACCTGCAGCGCCGAGCGCGTGGCGCGTGG 840
Db |||||
866 ATGTGACAGACACAGGTTCCAGGAGTACCTGCAGCGCCGAGCGCGTGGCGCGTGG 925
QY 841 GGCAGCGTCTATCTACCCGCTCCATGAAGAGGCGCATCGCCACGSCCATGAAGGTGGCCAG 900
Db |||||
926 GGCAGGCTCATCTACCCGCTCCATGAAGAGGCGCATCGCCACGSCCATGAAGGTGGCCAG 985
QY 901 GACACGTGGAGCTCGCAAGAACAGCTTTGAGCTCTACGCGGCTGACTTCGTCCTGGG 960
Db |||||
986 GACCACTGGAGCTCGCAAGAACAGCTTTGAGCTCTACGCGGCTGACTTCGTCCTGGG 1045
QY 961 AGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCGCTCAG 1020
Db |||||
1046 AGGACTTCAGGCGCTGGCTGATCGAGATCAATTCAGGCCCCACCATGACCCGCTCAG 1105
QY 1021 CCGGTCACGCGCCAGCTGTGTGCACAGGTGCAGGAGCACCATCAAGGTGGCGGTGGAC 1080
Db |||||
1106 CCGGTCACGCGCCAGCTGTGTGCACAGGTGCAGGAGCACCATCAAGGTGGCGGTGGAC 1165
QY 1081 CCGAGCTGTACATCGGCAACTTCGAGCTCCTGTGGAGCAGCCGCTGTTGAGCGCCC 1140
Db |||||
1166 CCGAGCTGTACATCGGCAACTTCGAGCTCCTGTGGAGCAGCCGCTGTTGAGCGCCC 1225
QY 1141 CCATTACGCGGTCGACCTCTGCTGTGGCGGGCTCAGTGTGAGGAGCGCAGGAGGAG 1200
Db |||||
1226 CCATTACGCGGTCGACCTCTGCTGTGGCGGGCTCAGTGTGAGGAGCGCAGGAGGAG 1285
QY 1201 GTGCTCCGCTGTGCAACTCTCAAGGCTCGGCTGCTGTGTGAGCGCGAGCGCTGAAG 1260
Db |||||
1286 GTGCTCCGCTGTGCAACTCTCAAGGCTCGGCTGCTGTGTGAGCGCGAGCGCTGAAG 1345
QY 1261 GCACGGGGCCCTCGGCCATGCCAGACCTTGCCAGGGACCCCCATCACAGCTCTCCAG 1320
Db |||||
1346 GCACGGGGCCCTCGGCCATGCCAGACCTTGCCAGGGACCCCCATCACAGCTCTCCAG 1405
QY 1321 CGGGACTTGGGACTGAAGAGAGGGGCTCCCTCGCTTGTGTCGACCTTTAAGG 1380
Db |||||
1406 CGGGACTTGGGACTGAAGAGAGGGGCTCCCTCGCTTGTGTCGACCTTTAAGG 1465
QY 1381 GCGGCGCAGCAGCGGTGGAGCGGCACAGCCGCCGACCAAGCTGTGGGAAGGTG 1440
Db |||||
1466 GGGGCGCAGCAGCGGTGGAGCGGCACAGCCGCCGACCAAGCTGTGGGAAGGTG 1525
QY 1441 GAGCTCCGGGCTGCCCTGTGCACAGTGGACAGTFCAGGCGCCCAACACCGGTGTC 1500
Db |||||
1526 GAGCTCCGGGCTGCCCTGTGCACAGTGGACAGTFCAGGCGCCCAACACCGGTGTC 1585
QY 1501 GTAGCCAGCCGCGCAAAAGCTGGGATCCAAACCACTAAATGCGCACCCGCTGGAGCT 1560
Db |||||
1586 GTAGCCAGCCGCGCAAAAGCTGGGATCCAAACCACTAAATGAGCACCCGCTGGAGCT 1645
QY 1561 GTGCTCGGGGCTGAAGACAGCAGAGGCGGCTCGCTCCGCGCCGCGAGGAAAGGT 1620
Db |||||
1646 GTGCTCGGAGCTGAAGACAGCAGAGGCGGCTCGCTCCGCGCCGCGAGGAAAGGT 1705
QY 1621 TCA 1623
Db |||
1706 TCA 1708
```

RESULT 6

ADJ93359

ID ADJ93359 standard; cDNA; 726 BP.

AC ADJ93359;

XX ADJ93359;

DT 06-MAY-2004 (first entry)

XX

Human BGS-42 cDNA sequence SeqID3.

testis-specific tubulin tyrosine-ligase-like polypeptide;
BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
brain cancer; liver cancer; proliferative condition; testis; lung;
small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.

Homo sapiens.

Key	Location/Qualifiers
CDS	1..726
FT	/*tag= a
FT	/product= "Human BGS-42 protein"
FT	/partial
FT	/note= "No start or stop codon"

WO2004005487-A2.

15-JAN-2004.

09-JUL-2003; 2003WO-US021605.

09-JUL-2002; 2002US-0394725P.

(BRIM) BRISTOL-MYERS SQUIBB CO.

Feder JN, Wu S, Nelson TC;

WPI; 2004-099381/10.

New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, e.g.
useful for preventing, treating or ameliorating a medical condition, e.g.
aberrant cellular proliferation, reproductive disorders or testicular
disorders.

Disclosure; SEQ ID NO 3; 343pp; English.

This invention relates to a novel testis-specific tubulin tyrosine-ligase
-like polypeptide, designated the BGS-42 polypeptide. The invention may
be useful for the development of compounds with a cytostatic, respiratory
-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
immunosuppressive, antiseborrheic or dermatological activity acting as
tyrosine ligase modulators. In addition, the disclosed sequences may be
useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
used for diagnosing a pathological condition or a susceptibility to a
pathological condition in a subject, and for preventing, treating or
ameliorating a medical condition, such as a disorder related to aberrant
tubulin ligase activity, a disorder related to aberrant tubulin-
carboxypeptidase activity, aberrant cellular proliferation, reproductive
disorders, testicular disorders, testicular cancer, pulmonary disorders,
lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
neural disorders, brain cancer, liver cancer, or proliferative condition
of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
polypeptide, polynucleotide, or their modulators are also useful for
treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
-42 polypeptide can be used as a preventive agent for immunological
disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
disease or scleroderma. The antibodies may be used to purify, detect and
target the BGS-42 polypeptides. The present sequence is that of a partial
cDNA which encodes the human BGS-42 protein of the invention.

XX
SQ Sequence 726 BP; 157 A; 227 C; 214 G; 128 T; 0 U; 0 Other;
Query Match 44.7%; Score 726; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 4.7e-145;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 397 ATTGACGGCTCCGGACATCTGGATTATAAAGCCGCGCCCAAGTCCCGGGGCCGAGAC 456
DB 1 ATTGACGGCTCCGGACATCTGGATTATAAAGCCGCGCCCAAGTCCCGGGGCCGAGAC 60
QY 457 ATAGTGTGCATGACCGCTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACCTCTTT 516
DB 61 ATAGTGTGCATGACCGCTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACCTCTTT 120
QY 517 TCAGGAGCAACAAAGTGGGTGTCAGAAATGATCATCGAGACGCGCTGCTCATCTGTGAC 576
DB 121 TCAGGAGCAACAAAGTGGGTGTCAGAAATGATCATCGAGACGCGCTGCTCATCTGTGAC 180
QY 577 ACCAAGTTCCACATCAGACAGTGTCTCTGTCACGAGCTGGAAACCCCTGACCATCTGG 636
DB 181 ACCAAGTTCCACATCAGACAGTGTCTCTGTCACGAGCTGGAAACCCCTGACCATCTGG 240
QY 637 TTCTACAAGGAGAGTACTTTCGCGGTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGAC 696
DB 241 TTCTACAAGGAGAGTACTTTCGCGGTCTCAACTCAGCGCTTCTCCCTGGCAAGCTGGAC 300
QY 697 AGGCGCATCATCTGTGCAACAAAGCGCTCCAGAAATGATCTGAAGATGATGTGGGCCG 756
DB 301 AGGCGCATCATCTGTGCAACAAAGCGCTCCAGAAATGATCTGAAGATGATGTGGGCCG 360
QY 757 AGCCCTCTGCTGCGCCGACACATGTGACAGCAGCAGCTTCCAGGAGTACCTGCGAC 816
DB 361 AGCCCTCTGCTGCGCCGACACATGTGACAGCAGCAGCTTCCAGGAGTACCTGCGAC 420
QY 817 CGCCAGGCGCTGCGCGCTGTGGGCGAGCTCATCTACCCCTCCATGAAGAAGGCCATC 876
DB 421 CGCCAGGCGCTGCGCGCTGTGGGCGAGCTCATCTACCCCTCCATGAAGAAGGCCATC 480
QY 877 GCCCAAGCATGAAGTGGCCCAAGGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 936
DB 481 GCCCAAGCATGAAGTGGCCCAAGGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTC 540
QY 937 TAGCGGGCTGACTTCGCTCTTGGAGGGAGCTTCAGGCGCTGGCTGATCGAGATCAATTCC 996
DB 541 TAGCGGGCTGACTTCGCTCTTGGAGGGAGCTTCAGGCGCTGGCTGATCGAGATCAATTCC 600
QY 997 AGCCCAACCATGACACCGCTCCAGCGCTGTCAGCGCCAGCTGTGTGACAGGTGCAAGGAG 1056
DB 601 AGCCCAACCATGACACCGCTCCAGCGCTGTCAGCGCCAGCTGTGTGACAGGTGCAAGGAG 660
QY 1057 GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTTCTGTGG 1116
DB 661 GACACCATCAAGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTTCTGTGG 720
QY 1117 AGGCAG 1122
DB 721 AGGCAG 726

RESULT 7
ADQ17814/c
ID ADQ17814 standard; DNA; 101270 BP.
XX
AC ADQ17814;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 631.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX
ds.

OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PP 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
XX Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
XX Early detection of soft tissue sarcoma comprises determining expression
of a gene in a first soft tissue sample and a normal soft tissue sample
and comparing the gene expression, also useful in treating soft tissue
sarcoma.
XX
XX Example 2; SEQ ID NO 631; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
which comprises obtaining a first soft tissue sample from an individual
and a normal soft tissue sample from the same or different individual,
determining the expression of a gene in both samples and comparing the
expression of the gene in both soft tissue samples, where a higher level
of protein expression in the first soft tissue sample indicates the
presence of soft tissue sarcoma. The method of the invention has
cytostatic applications and may be useful for detecting soft tissue
sarcoma, possibly via gene therapy or vaccine production. The nucleic
acid sequences may be useful in diagnostic and screening applications.
The current sequence is that of a human soft tissue sarcoma-upregulated
DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 101270 BP; 24151 A; 28044 C; 28100 G; 20975 T; 0 U; 0 Other;
Query Match 30.9%; Score 501; DB 12; Length 101270;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1120 CAGCCGGTGTGAGCGCGCCCATTCAGCGGTTCGACCTCTGCTGGCGGGCTCAGT 1179
DB 43637 CAGCCGGTGTGAGCGCGCCCATTCAGCGGTTCGACCTCTGCTGGCGGGCTCAGT 43578
QY 1180 GTGAGGAGAGCCAGGAGGAGGAGTGTGCTGCTGCAACCTCAAGGCTCGGCTCGCTG 1239
DB 43577 GTGAGGAGAGCCAGGAGGAGGAGTGTGCTGCTGCAACCTCAAGGCTCGGCTCGCTG 43518
QY 1240 TTGAGCGCGCAGCGCTGAAGGACAGCGGGCCCTCGGCCATGCGCAGACCTGCCAGGGA 1299
DB 43517 TTGAGCGCGCAGCGCTGAAGGACAGCGGGCCCTCGGCCATGCGCAGACCTGCCAGGGA 43458
QY 1300 CCCCACATCAGCTCTCCAGCGGAGCTTGGAGCTTGAAGAGAGAGAGAGGCTCCCGCTG 1359
DB 43457 CCCCACATCAGCTCTCCAGCGGAGCTTGGAGCTTGAAGAGAGAGAGAGGCTCCCGCTG 43398
QY 1360 GCCTTGTGTCACCTTTAAGGGGGGCGAGCGAGCGGTGGAGCCGACAGCCACCCCGC 1419
DB 43397 GCCTTGTGTCACCTTTAAGGGGGGCGAGCGAGCGGTGGAGCCGACAGCCACCCCGC 43338
QY 1420 ACCAAAGCTGTGGGAAGGTGGAGCTCCCGGGCTGCGCCCTGTGCGCACGCTGGACAGT 1479
DB 43337 ACCAAAGCTGTGGGAAGGTGGAGCTCCCGGGCTGCGCCCTGTGCGCACGCTGGACAGT 43278
QY 1480 GCCCCAAACACCGGTGTCCCGTAGCCCGCCGCAAGAGCTGGGATCAAAACAGCTA 1539
DB 43277 GCCCCAAACACCGGTGTCCCGTAGCCCGCCGCAAGAGCTGGGATCAAAACAGCTA 43218
QY 1540 AATGCGACCCCTGGAGCTGTGCTGCGGGGCTTGAAGACAGCAGAGGCGGCTGCGT 1599

Db 43217 AATGCGACCCGCTGGAGCCTGTGCTGCGGGCCTGAACAGCAGCAGAGGCGCGCTGCGT 43158
QY 1600 CCGCGCGCCGAGGAAAGGT 1620
Db |||||
43157 CCGCGCGCCGAGGAAAGGT 43137
RESULT 8
AAS99894
ID AAS99894 standard; cDNA; 2380 BP.
XX |||||
AC AAS99894;
XX |||||
DT 12-MAR-2002 (first entry)
XX
DE Polynucleotide encoding human cytoskeleton-associated protein (CYSKP) #5.
XX
KW Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;
KW cell proliferative disorder; inflammatory disorder; prion disease;
KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
KW neurological disorder; cell motility disorder; reproductive disorder;
KW spinal cord disease; central nervous system disorder; mental disorder;
KW gene therapy; cancer.
XX
OS Homo sapiens.
XX
XX WO200185942-A2.
XX
XX 15-NOV-2001.
XX
XX 03-MAY-2001; 2001WO-US014355.
XX
XX 05-MAY-2000; 2000US-0201960P.
PR 08-MAY-2000; 2000US-0202729P.
PR 05-JUN-2000; 2000US-0209705P.
PR 07-JUN-2000; 2000US-0210149P.
PR 21-JUN-2000; 2000US-02113215P.
XX
XX (INCYTE GENOMICS INC.
XX
XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;
PI Azinzaiz Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;
PI Policky JL;
XX
XX WPI; 2002-062248/08.
DR P-PSDB; AAU74334.
XX
XX New cytoskeleton-associated proteins and polynucleotides, useful for
PT diagnosing, preventing and treating cell proliferative, autoimmune,
PT inflammatory, neurological, cell motility, reproductive and muscle
PT disorders.
XX
XX Claim 5; Page 171; 194pp; English.
XX
XX The invention relates to human cytoskeleton-associated polypeptides
CC (CYSKP) and their associated polynucleotide sequences. The sequences are
CC useful in the treatment of disorders associated with overexpression or
CC underexpression of CYSKP in a patient. The disorders include cell
CC proliferative disorders (such as cancer, actinic keratosis,
CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),
CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,
CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus
CC and anaemia), vesicle trafficking disorders (such as
CC hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),
CC gastrointestinal disorders, prion diseases, neurological disorders (such
CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,
CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis
CC and other motor neuron disorders), cell motility disorders, reproductive
CC disorders (such as endometriosis and polycystic ovary syndrome), muscle
CC disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,
CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord
CC diseases, central nervous system disorders (such as Down syndrome and
CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

CC Sequences AAS99890-AAS99923 represent cDNA molecules encoding human CYSKP
CC of the invention
XX
SQ Sequence 2380 BP; 580 A; 685 C; 670 G; 445 T; 0 U; 0 Other;
Query Match 26.7%; Score 432.8; DB 6; Length 2380;
Best Local Similarity 59.1%; Pred. No. 2e-82;
Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;
QY 108 CAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGTCTAG-GGGCCCTCCGGGGCAGCTTG 166
Db |||||
361 CAGGAGACAGCAGCCCAAGAAACAGGAGAAAAACCCAGTGTGGTGTCCCCAGAGTTTG 420
QY 167 TGGACATCGCGTGCACAGGTGTGCCAGGCTACTCTGGGCGAGCTGAGGACATGAGGACATCG 226
Db |||||
421 TGGATGAAGCTCTGTGTGCTGCGAGGAGTACTTTAGCAACTTGGCCCAATGGGACATCG 480
QY 227 ACACGTGACGAGATCCCGTGGAGGACCTCACTGAGSCCGAGTGGGAGGACCTGACCCAGC 286
Db |||||
481 ACAAGGACCTGGAGGCCCGCTGTACCTACACCCGAGGGCTGGTCCCTCTTCCTCCAGC 540
QY 287 AGTACTACTCCCTCGTTATGCGGATGCTTTCATCTCCAATTCAGAAATTAATTTTCG 346
Db |||||
541 GCTACTCAAGTGTGCCAGGAGGCGCAACTCAGGCACCTCGACACTCAGTCCAGC 600
QY 347 AGTGCCAGGCTCTGCTGAATAGAAATCACTGTGTGAACCTCTCAGACGACATTTAGCGGC 406
Db |||||
601 GCTGTGAGGACATCTCTGACAGAGCTGCAGGCCGTGGTACCCAGATAGACATGGAGGG 660
QY 407 TCCGGAACATCTGGATTATAAGCCGCGGCCCAAGTCCCGGGCCGAGACATAGTGTGA 466
Db |||||
661 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCCAAGTCCCGTGAGGAGCATCATGTGA 720
QY 467 TGGACCGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGACA 526
Db |||||
721 TGGACCACTGGAGGAGATGCTGAAGTGTGAACGGCAACCCCGTGGTGAAGAGCG 780
QY 527 ACAAGTGGTGTCCAGAAATACATCGAGACGCGCTGTCTATCTGTGACACCAAGTTTCG 586
Db |||||
781 GCAAGTGGTGTGCAGAAATATATTGAGCGGCCCTCTCTCATCTTTGGCACCAGTTTG 840
QY 587 ACATCAGACAGTGGTTCCTCGTCGCGACTGCAACCCCTGACCATCTGCTTCTACAAG 646
Db |||||
841 ACCTCAGACAGTGGTTCCTGGTAACTGACTGGAACCCCACTTACCGTGTGGTCTACCGCG 900
QY 647 AGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGACAGCGCCATCC 706
Db |||||
901 ACAGCTATATCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCTGAGCAACTCAGTGC 960
QY 707 ACTGTGCAACAAACCGCGTCCAGAGTACCTGAAGATGATGTGGGCCGAGCCCTCTGC 766
Db |||||
961 ACCTGTGCAACAACTCCATCCAGAAAGCACCTGGAGAACTCATGCCATCGGCATCCACTGC 1020
QY 767 TGC CGC CAC AACA CAT GTG GAC CAC CAG GTT CC AGG AT TCC AGG AT TAC CTG AG CGC CAG GGC 826
Db |||||
1021 TTCCGCGCAGACAAACATGTGGTCTAGCAGAGGTTCAGGSCCCACCTGCGAGGAGATGGTG 1080
QY 827 GTGGCGCGCTGTGGGCGCAGCGTCATCTACCCGTCCATGAAGAGGCCATCGCCACGCCA 886
Db |||||
1081 CCCCAAATGTTGGTCCACCATCATCTGCTGGCATGAAGGATGCTGTGATCCACGCAC 1140
QY 887 TGAAGGTGGCCCGGACCAACGTGGAGCCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTG 946
Db |||||
1141 TTCAAGACCTCCCGAGCACCGTGCAGTGTGCGAAGGCCAGCTTTGAGCTCTATGGCGCTG 1200
QY 947 ACTTGTCTTTGGGAGGAGCTTTCAGGCCCTGCTGATCGAGATCAATTTCCAGCCCCACCA 1006
Db |||||
1201 ACTTGTGTTCGGGGAGGACTTCCAGCCCTGGCTGATTCAGATCAACGCCAGCCCCACCA 1260
QY 1007 TGCACCCGCTCCACGCGCGTCAAGGCCCTGCTGTGACAGAGTGCAGAGGAGCACCATCA 1066
Db |||||
1261 TGGCACCTCCACAGCAGCATCTGCGCCGCTCTGTGCTGGCGTGCAGAGCTGACACCTGC 1320


```
QY 1067 AGTGGCC-----GTGACCGCAGCTGTGACATCGGCAACTTCGAGCTCCTGT 1114
D 1321 CGTGTGATTCAGCGATGCTGACCGCAACTGTGACAGGAGCTTTGAGCTCACT 1380
QY 1115 GGAGGAGCGGTGTTGAGCGCGCCCAATTCAGCGGGTCCGACTCTGGTGTGCGGGCG 1174
D 1381 ATAAGCAGCTGTGTGAGGTCTCAATATGTGGGCATCCGGCTCTCTGTAGAGGCT 1440
QY 1175 TCAGTGTGAGGAGCCAGAGGAGGTGTGCGGCTGTGCACTCAAGGCTTCGGCT 1234
D 1441 TCACCATCAAGAGCCATCGCATGTGTATCGCGGATGGGGTCCGCCACGATCC 1500
QY 1235 CGCTGTGGACGCGCAGCGCTGAAGGACAGCGGCCCTCGGCCATGCC-AGACCTGTGC 1293
D 1501 CTCTGTGACCCAGCAGGCTCTGGGGAAGCAAGGACTCGGGATCCCTACCCACAGT 1560
QY 1294 CAGGAGCCCCATCACCAGCTCTCCAGCGGACTTGGGACTGAAGGAAGAGGGGCTC 1353
D 1561 CAGCTTCTAGGAAAGGCACTGGGSCCAGGAGCTTGGGGCACAGTGAGAGCCAGTCTCCA 1620
QY 1354 CCCCTGGCTTGTGGCACCTTAAGGGGGGAGCCGAGAGCGGTGAGCGGCACAGCCC 1413
D 1621 CTGCGCACCTCTCAGCCCGGAAAGGGAAGCAAGGCAAGGCAAGGCGCAAGGCCACAGCCC 1680
QY 1414 ACCCGACCAAGCTGTGCGGAGGTGGAGCTCCCGC 1451
D 1681 TGGTCTGCCCAATCTCTGGAGTGGATGCCCCCAGC 1718

RESULT 9
AAH16735
ID AAH16735 standard; cDNA; 2326 BP.
XX
AC AAH16735;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15920.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
XX
PR 27-AUG-1999; 99JP-00300253.
XX
PR 11-JAN-2000; 2000JP-00118776.
XX
PR 02-MAY-2000; 2000JP-00183767.
XX
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isozaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 15920; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
```

```
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH58893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2326 BP; 524 A; 701 C; 642 G; 459 T; 0 U; 0 Other;
```

```
Query Match 26.6%; Score 431.6; DB 4; Length 2326;
Best Local Similarity 63.5%; Pred. No. 3.5e-82;
Matches 698; Conservative 0; Mismatches 389; Indels 13; Gaps 2;
```

```
QY 108 CAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCTCCCGGGCAGCTTG 166
D 620 CAGGAGACAGCAGCCCCAAGAAAACAGAGAAAAACCCAGTGTGGTGTCTCCCAAGTGTG 679
QY 167 TGGACATCGCGTCAAGGTGTGCGAGGCTTACCTGGGGCAGCTGGGACATGAGGACATCG 226
D 680 TGGATGAAGCTCTGTGTGCTGCGAGGATACCTTAGCAACTTGGCCCAATGAGACATCG 739
QY 227 ACACGTGACAGATGCGGTGGAGGACCTCACTGAGGCGGAGTGGGAGGACCTGACCCAGC 286
D 740 ACAAGGACCTGGAGGCCCCGCTGCTACCTACCCCGAGGGCTGTGCTCTCTCTCCAGC 799
QY 287 AGTACTACTCCCTGCTTCATGGCGATGCTTTCATCTCCAATTCAGAAATTAATCTTTGCG 346
D 800 GCTACTACCAAGTGGTCCAGAAAGGGGCGAGAACTCAGGCACTCGACACATCAGGTCCAGC 859
QY 347 AGTGCCAGGCTCTGCTGAATAGAATACGCTCTGTGAACCTCTAGAGCGACATTTGACGGGC 406
D 860 GCTGTGAGGACATCTCTGCGAGCAGCTGAGGCCGTGTGTACCCAGATAGACATGGAAGGG 919
QY 407 TCCGGAACATCTGGATTATAAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGCA 466
D 920 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGCGAGGAGCATCATGTGCA 979
QY 467 TGGACGCTGTGGAGGAGATCTTGAGCTGCGACCTGAGACCAACCTCTTTCAGGGGACA 526
D 980 TGGACCACCTGGAGGAGATGCTGAAGCTGGTGAACCGGCAACCCCGTGTGTATGAAGGAGC 1039
QY 527 ACAAGTGGGTGGTCCAGAAAGTACATCGAGACGCGCTGTCTCATCTGTGACACCAAGTTTC 586
D 1040 GCAAGTGGGTGGTGCAGAAAGTATATTGAGCGGCCCTCTCTCATCTTTGGCACCAGTTTG 1099
QY 587 ACATCAGACAGTGGTTCCCTCGTCACGAGCTGGAAACCCCTTGACCATCTGTTCTTACAAGG 646
D 1100 ACCTCAGACAGTGGTTCTCTGTTAACTGACTTGAACCCCACTTACCGTGTGTCTTACCGCG 1159
QY 647 AGATTACTTGGCGTTCTCAACTCAGCGTTCTTCCCTGTGGAACAAGCTGGACAGCGCCATCC 706
D 1160 ACAGCTATATCCCTTTTCCACGAGCCCTTCTCTCCCTGAAGAAGACCTTGGAACAATCAGTGC 1219
QY 707 ACCTGTGCAACACGCGCTCCAGAAAGTACCTGAAGATGATGTGGGCGCGAGCCCTCTGC 766
D 1220 ACCTGTGCAACAACTCCATCCAAAGACCTGAGAACTCATGCGATCGGCATCTCACTGC 1279
QY 767 TGCCCCGACACAACATGTGGACACGAGCCAGGTTTCAGGAGTACCTGCGAGCGGCGGCC 826
D 1280 TTCCGCCAGACAACATGTGGTCTTAGCCAGAGGTTCCAGGGCCCACTCTGCAGGATGGGTG 1339
```


QY 827 GTGGCCCGGTGTGGGCGAGCGTCACTACCCGTCCATGAAGAGGCATCGGCCACGCCA 886
Db 1340 CCCAAATGTTGGTCCACCATCATGCTGCTGGCATGAAGATGCTGTGATCCAGCAC 1399
QY 887 TGAAGTGGCCAGGACCAAGTGGAGCCTCGCAAGACAGCTTGAAGCTTACGGGGCTG 946
Db 1400 TTGAGACCTCCCAAGGACACCGTGCAGTGTGGAAGGCCAGCTTGAAGCTTATGGGGCTG 1459
QY 947 ACTTCGTCTCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCCACCA 1006
Db 1460 ACTTCGTGTGGGAGGAGCTTCAGGCCCTGGCTGATTCAGATCAAGCCAGCCACCA 1519
QY 1007 TGCACCCGTCCAGCCGCTCAGGCCCGCCAGCTGTGTGCAAGGTGAGGAGGACACCATCA 1066
Db 1520 TGGCACCCCTCCACAGCAGTCACTGCCCGGCTGTGTCTGGCGTGCAGCTGACACCCCTGC 1579
QY 1067 AGTGGCC-----GTGGACCGCAGCTGTGACATCGCAACTTCGAGCTCCTCT 1114
Db 1580 GGTGTGTCATTACCGGATGCTGGACCGCACTGTGACACAGAGCCTTTGAGCTCATCT 1639
QY 1115 GGAGGCAGCGGTGTGTGAGCCGCCCCCATTCAGCGGGTCCGACCTCTCGCTGGCGGGCG 1174
Db 1640 ATAGCAGCGCTGCTGTGAGGTGCTCAATATGTGGGCATCGGCTCCTGTAGAGGGCT 1699
QY 1175 TCAGTGTGAGGAGCCAGG 1194
Db 1700 TCACCATCAAGAAGCCCATG 1719

RESULT 10
ACH91699/c
ID ACH91699 standard; DNA; 490 BP.
XX ACH91699;
AC ACH91699;
XX ACH91699;
DT 29-JUL-2004 (first entry)
XX Human genome derived single exon probe #24894.
DE Human; probe; ss; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX Homo sapiens.
OS US2003194704-A1.
PN 16-OCT-2003.
XX 03-APR-2002; 2002US-00029386.
XX 03-APR-2002; 2002US-00029386.
XX (PENN/) PENN S G.
XX (RANK/) RANK D R.
XX (HANZ/) HANZEL D K.
XX Penn SG, Rank DR, Hanzel DK;
XX WPI; 2004-119264/12.
XX
XX New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
PT surveying tissues.
XX Claim 1; SEQ ID NO 24894; 80pp; English.
XX
XX The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule

CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subcription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterising
CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ Sequence 490 BP; 77 A; 150 C; 168 G; 95 T; 0 U; 0 Other;
Query Match 26.3%; Score 427; DB 12; Length 490;
Best Local Similarity 100.0%; Pred. No. 2.4e-81;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 696 CAGCGCCATCCACCTGTGTGCAACAACCGCGTCCAGAGTACCTGGAAGATGATGTGGCGG 755
Db 427 CAGCGCCATCCACCTGTGTGCAACAACCGCGTCCAGAGTACCTGGAAGATGATGTGGCGG 368
QY 756 CAGCGCCCTGTGTGTGCGGCACACAACATGTGGACCAAGTCCAGGTTCCAGGATCTGCA 815
Db 367 CAGCGCCCTGTGTGTGCGGCACACAACATGTGGACCAAGTCCAGGTTCCAGGATCTGCA 308
QY 816 GCGCCAGGCGCGTGGCGCGTGTGGCGAGCGCTCATCTACCGCTCCATGAAGAAGGCCAT 875
Db 307 GCGCCAGGCGCGTGGCGCGTGTGGCGAGCGCTCATCTACCGCTCCATGAAGAAGGCCAT 248
QY 876 CCGCCCAAGGAGTGGCCAGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCT 935
Db 247 CCGCCCAAGGAGTGGCCAGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCT 188
QY 936 CTACGGGGCTGACTTCTGCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTC 995
Db 187 CTACGGGGCTGACTTCTGCTTGGGAGGAGCTTCAGGCCCTGGCTGATCGAGATCAATTC 128
QY 996 CAGCCCCCACCATGCACCCGCTCCACCGCCGCTCACCGCCAGCTGTGTGCAAGTGCAGGA 1055
Db 127 CAGCCCCCACCATGCACCCGCTCCACCGCCGCTCACCGCCAGCTGTGTGCAAGTGCAGGA 68
QY 1056 GGACACCATCAAGGTGGCGGTGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTG 1115
Db 67 GGACACCATCAAGGTGGCGGTGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTG 8
QY 1116 GAGGCAG 1122
Db 7 GAGGCAG 1
RESULT 11
ADS73168
ID ADS73168 standard; cDNA; 2538 BP.
XX

AC ACN37881;
XX 18-NOV-2004 (first entry)
XX Tumour-associated antigenic target (TAT) cDNA DNA324273, SEQ ID NO:1057.
DE Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic; gene; ss.
XX Homo sapiens.
XX WO2004030615-A2.
PN 15-APR-2004.
XX 29-SEP-2003; 2003WO-US028547.
XX 02-OCT-2002; 2002US-0414971P.
PR (GETH) GENENTECH INC.
XX Wu TD, Zhang Z, Zhou Y;
XX WPI: 2004-347921/32.
DR P-FSDB; ABM80420.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 1; SEQ ID NO 1057; 7273pp; English.
XX
CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
SQ Sequence 2553 BP; 575 A; 716 C; 733 G; 529 T; 0 U; 0 Other;
Query Match 26.0%; Score 421.4; DB 13; Length 2553;
Best Local Similarity 63.8%; Pred. No. 5.3e-80;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 80 GGGAGAGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGTGTGAAATGCTGAGGCAA 139
Db 754 GGGATCGTGACAGGCGCCCTCCCTATGTGTCAGGAGACAGCAGCCCAAGAAACAGGAGAA 813
QY 140 AGCTCAG-GGGCTCCCGGGGAGCTTGTGGACATCGCGTGAAGTGTGCCAGGCGCTAC 198
Db 814 AACCCAGTGTGGTGTCCCGAGAGTTGTGGATGAAGCTCTGTGTGCGTGGAGGAGTAC 873

QY 199 CTGGGGCAGCTGGAGCATGAGGACATCGACACGTCAGCAGATGCCGTGGAGAGCTCACT 258
Db 874 CTTAGCAACTTGGCCCAATGACATCGACAGGACCTGGAGGCCCGCTGTACTCACC 933
QY 259 GAGGCGGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGCGCATCTTC 318
Db 934 CCGGAGGGCTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGGTCCACGAAGGGGAGAA 993
QY 319 ATCTCCAAATTAAGAAATTAATCTTTTCGAGTGCAGGCTCTCTGCTGAATAGAATCACTCT 378
Db 994 CTCAGGCACCTCGACACTCAGCTCAGCGCTGTGAGGACATCTCGCAGCAGTGCAGGCC 1053
QY 379 GTGAACCTTCAGACGGACATTCAGCGGCTCCGAAACATCTCGAATATAAAGCCCGCGGC 438
Db 1054 GTGGTACCCAGATAGACATGGAAGGGGATCGCAACATCTGGATCGTGAAGCAGAGCC 1113
QY 439 AGTCCCGGGGCGGAGACATAGTGTGCATGAGCCGCTGTGGAGGAGATCTCGAGCTGGCA 498
Db 1114 AAGTCCCGGAGCAGGAGCATCATGTGCATGGACCACTCTGGAGGAGATGCTGAAGCTGGTG 1173
QY 499 GCTGCAGACCACTCTTTCCAGGGAACAAGTGGTGGTCCAGAAATACATCGAGAGC 558
Db 1174 AAGGCAACCCCGTGGTGTGATGAAGACGGCAAGTGGTGCAGAAATATATTGAGCGG 1233
QY 559 CCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGTTCCTCTCTCGGACTGG 618
Db 1234 CCCCTCCTCATCTTTGGCACCAGTTTGACCTCAGACAGTGTTCCTCTGTAACCTGACTGG 1293
QY 619 AACCCCTGACCATCTGGTTCTACAAGGAGAGTACTTTGCGGTTCTCACTCAGCGCTTC 678
Db 1294 AACCCACACTTACCGTGTGGTTTACCGGACACGCTATATCCGCTTTTCCACGAGCCCTTC 1353
QY 679 TCCCTGGAAGCTGGAAGCGCCCATCCACTGTGCAACAACGCGGTCCAGAAATACCTG 738
Db 1354 TCCCTGAAGACCTGGAACACTCAGTGCACCTGTGCAACAACCTCCATCCAGAAACACTG 1413
QY 739 AAGAATGATGTGGCGCGCAGCCCCCTGTGTGCCCGCACAAACATGTGGAACGACCAAG 798
Db 1414 GAGAACTCATGCGCATCGGCATCCACTGTCTCCGCGCAGACAACATGTGTCTAGCCAGAG 1473
QY 799 TTCGAGGAGTACCTCAGCGCGCGCGCGGTGGCGCGGTGTGGGCGAGCGTCACTACCG 858
Db 1474 TTCCAGGCGCCACTCAGGAGATGGTGGCGCCCAATGTGTGTCCACCATCATCTGTGCT 1533
QY 859 TCCATGAAGAAGCGCATGCCCAACGCGCATGAAGTGGCGCAGGACCACTGGAGCGCTCG 918
Db 1534 GGCATGAAGGATGCTGTGATCCAGCACTTCAGACCTCCAGGACACCTGCGAGTCTCG 1593
QY 919 AAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGTCTTTGGAGGGAATTCAGGCGCTGG 978
Db 1594 AAGGCCAGCTTTGAGCTCTATGGCGCTGACTTCGTGTTCGGGAGGAGACTTCCAGCGCTGG 1653
QY 979 CTGATCGAGATCAATTTCCAGCGCCCACTGACACCGCTCCAGCGGTCCAGCGCCAGCTG 1038
Db 1654 CTGATGAGATCAAGCCAGCGCCCACTGAGTGGACCCCTCCAGAGAGTCACTGCGCGCTC 1713
QY 1039 TGTGCACAGGTGTCAGGAGGACACCATCAAGGTGTGGCC-----GTGGACCGCAGC 1086
Db 1714 TGTGCTGGCTGCAAGCTGACACCTCGCGGTGGTCAATTGACCGGATGCTGCGACCGCAAC 1773
QY 1087 TGTGATCGGCAACTTTCGAGCTCTGTGAGGAGGAGCGCGGT 1127
Db 1774 TGTGACACAGGAGCGCTTTGAGCTCATCTATAGCAGCGCGT 1814
RESULT 13
AAC77214
ID AAC77214 standard; cDNA; 2881 BP.
XX
AC AAC77214;
XX
DT 08-FEB-2001 (first entry)
XX

XX AAI58606;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX Human polynucleotide SEQ ID NO 809.
DE
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
OS
XX WO200153312-A1.
PN
XX
XX 26-JUL-2001.
PD
XX 26-DEC-2000; 2000WO-US034263.
PF
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Zhou J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR P-PSDB; AAM39450.
DR
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Claim 1; SEQ ID NO 809; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 3001 BP; 662 A; 854 C; 846 G; 639 T; 0 U; 0 Other;

Query Match 26.0%; Score 421.4; DB 4; Length 3001;
Best Local Similarity 63.8%; Pred. No. 5.5e-80;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;

QY 80 GGGAGGAGCGGGAGCAGCACCTTGTGGACATCGCGTGAAGGTGTGCCAGGCGCTAC 139
DB 1223 GGGATCGTGACAGGCCCTCCCTTGTGTCAGAGACAGCAGCCCAAGAAACAGGAGAAA 1282
QY 140 AGCTCAG-GGGGCTCCCGGGCGAGCTTGTGGACATCGCGTGAAGGTGTGCCAGGCGCTAC 198

Db 1283 AACCCAGTGTGTGTGTCGCCAGAGTTTGTGGATGAAGCTCTGTGTGCTGCCAGGAGTAC 1342
QY 199 CTGGGGCAGCTGGAGCATGAGGACATCGACAGCTCAGCAGATGCGCTGGAGGAGCTTCACT 258
Db 1343 CTTAGCAACTTTGGCCCAATGACATCGACAAGGACTGGAGGCCCGCTGTGTACCTCACC 1402
QY 259 GAGGCGCAGTGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGCCGATGCTTTC 318
Db 1403 CCGGAGGCTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGTGTCCAGAAAGGGGAGAA 1462
QY 319 ATCTCAATTAAGAAATTAATTTTTCGAGTGCAGGCTCTGTCTGAATAGAAATCAGCTCT 378
Db 1463 CTCAGGCACCTCGACACTCAGGCTCAGGCTGTGTGAGACATCTCTGCAGCAGCTGCAGGCC 1522
QY 379 GTGAACCTCTCAGACGACATTCAGCGGCTCCGGAACATCTGGATTTAAAGCCCGCGGCC 438
Db 1523 GTGGTACCCCAATAGATGAAGGAGGATTCGCAACATCTGGATCTGGAAGCAGGAGCC 1582
QY 439 AAGTCCCGGGCGCGAGACATAGTGTGATGGACCGTGTGGAGGAGATCTCTGGAGCTGGCA 498
Db 1583 AAGTCCCGGGAGGAGGATCATGTGATGGACCACTTGGAGGAGATGCTGAAGCTGGTG 1642
QY 499 GGTGCAGACCACTCTTTTCCAGGGACAAAGTGGTGGTCCAGAAAGTACATCGAGACG 558
Db 1643 AACGGCAACCCCGTGTGTGATGAAGGACCGCAAGTGGTGGTGTGAGAAAGTATATTGAGCGG 1702
QY 559 CCGTCTCTATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCTCTCAGGACTGG 618
Db 1703 CCCCTCTCTATCTTTGGCACCAAGTTTGACCTCAGACAGTGGTTCCTGGTAACTGACTGG 1762
QY 619 AACCCCTCGACCATCTGGTTCTACAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTC 678
Db 1763 AACCCACTTACCGTGTGGTTCTACCGGACAGCTATATCGGCTTTTCCAGCGAGCCCTTC 1822
QY 679 TCCCTGGACAAGCTGACAGCGCCATCCACCTGTGCAACAAAGCGCTCCAGAAAGTACCTG 738
Db 1823 TCCCTGAAGAACCTGGCAAACTCAGTGCACTGTGCAAACTCCATCCAGAAAGCACCTG 1882
QY 739 AAGATGATGTGGCGCGCAGCCCCCTGCTGCCGACACAAACATGTGGACCGACACAGG 798
Db 1883 GAGAACTCATGCGCATCGGCATCCTGCTTCCGCCAGACAACATGTGGTCTAGCCAGAGG 1942
QY 799 TTTCCAGGAGTACTCTGAGCGCCAGGCGCGTGGCGCGCTGTGGGGCAGCGCTCATCTACCCG 858
Db 1943 TTTCCAGGCGCCACTGACAGGAGATGGTGGCCCCAAATGCTTGGTCCACCATCATCTGTCCT 2002
QY 859 TCCATGAAGAAAGCCATCCGCCACGCGCATGAAGGTGGGCCAGGACCAACGCTGGAGCCCTGC 918
Db 2003 GGCATGAAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTCCG 2062
QY 919 AAGAACAGCTTTGAGCTTACCGGGCTGACTTCGTCTTGGGAGGAGACTTCAGGCCCTTG 978
Db 2063 AAGGCCAGCTTTGAGCTTATGGCGCTGACTTCGTGTTTCGGGAGGAGACTTCCAGCCCTTG 2122
QY 979 CTGATCGAGATCAATTCACAGCCCCACCATGCAACCCCTCCAGCGCGCTCAGCGCCAGCTG 1038
Db 2123 CTGATTTGATTCACCGCAGCCCCCAGGATGGACCCCTCCACAGCAGTCACTACTGCCCGCTC 2182
QY 1039 TGTGCACAGGTGCAGGAGGACACCATCAAGGTGGCC-----GTGGACCGGAGC 1086
Db 2183 TGTGCTGGCGTGCAAAGCTGACACCCCTCGCGTGGTTCATTGACCGGATGCTGGACCGCAAC 2242
QY 1087 TGTGACATCGGCAACTTCGAGCTCTCTGTGGAGGAGCCCGGT 1127
Db 2243 TGTGACACAGGAGCCTTTGAGCTCATCTATAAGACGACCCGT 2283

RESULT 15
ADQ98824
ID ADQ98824 standard; cdna; 3001 BP.
XX
AC ADQ98824;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 24, 2005, 11:30:53 ; Search time 7533.96 Seconds
(without alignments)
10438.446 Million cell updates/sec

Title: US-10-615-659-1_COPY_153_1775
Perfect score: 1623
Sequence: 1 atggcattcagcattcctcaaa.....cgccggaggaagaggttca 1623

Scoring table: IDENTITY NUC
Gapex 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	817	50.3	817	6	CQ724907 Sequence
2	501	30.9	101270	9	HS355C18
3	432.8	26.7	2380	6	AX301197 Sequence
4	431.6	26.6	2326	6	BD158727 Primer fo
5	431.6	26.6	2326	6	AX881015 Sequence
6	431.6	26.6	2326	9	AK023960 Homo sapi
7	431.6	26.6	4238	9	HSMB05098
8	421.4	26.0	2553	9	HSMB00637
9	421.4	26.0	3001	6	AR339003 Sequence
10	418	25.8	2848	6	AX834642 Sequence
11	418	25.8	2848	9	AK097236 Homo sapi
12	380.2	23.4	1684	9	AF078842
13	366.4	22.6	1857	10	BC006830
14	283.6	17.5	163783	10	AC119959
15	283.6	17.5	201420	2	AC117700
16	277.2	17.1	205949	2	AC134940
17	277.2	17.1	218249	2	AC097425
18	231.4	14.3	5282	6	CQ842940
19	231.4	14.3	5282	9	AK125875 Homo sapi

C 20	224.6	13.8	3828	6	CQ850619	Sequence
C 21	224.6	13.8	3828	9	AK127786	Homo sapi
C 22	213	13.1	2979	6	CQ595968	Sequence
C 23	192.2	11.8	145435	2	AC026685	Homo sapi
C 24	192.2	11.8	161903	2	AC021996	Homo sapi
C 25	191.4	11.8	1560	9	BC009479	Homo sapi
C 26	190.6	11.7	1958	6	BD160681	Primer fo
C 27	190.6	11.7	1958	6	AX884044	Sequence
C 28	190.6	11.7	1958	9	AK024110	Homo sapi
C 29	190.6	11.7	155313	2	AC068315	Homo sapi
C 30	190.6	11.7	185067	9	AC022382	Homo sapi
C 31	190.6	11.7	189430	2	AC011610	Homo sapi
C 32	190.6	11.7	191834	2	AC026196	Homo sapi
C 33	190.6	11.7	197360	2	AC018829	Homo sapi
C 34	190.6	11.7	216311	9	AC018809	Homo sapi
C 35	181.2	11.2	224077	2	AC120010	Homo sapi
C 36	181.2	11.2	266888	2	AC137881	Mus muscu
C 37	174.2	10.7	2543	6	CQ595656	Sequence
C 38	146.4	9.0	220290	2	BX936393	Danio rer
C 39	145.8	9.0	161903	2	AC021996	Homo sapi
C 40	145.8	9.0	222994	2	AC016947	Homo sapi
C 41	144.8	8.9	4615	6	CQ595655	Sequence
C 42	144.8	8.9	76854	3	AC003052	Drosophil
C 43	144.8	8.9	86398	2	AC017785	Drosophil
C 44	144.8	8.9	167201	3	AC092397	Drosophil
C 45	144.8	8.9	259718	3	AE003614	Drosophil

ALIGNMENTS

RESULT 1	CQ724907	Sequence 10841 from Patent WO02068579.	817 bp	DNA	linear	PAT 03-FEB-2004
LOCUS	CQ724907					
DEFINITION	CQ724907					
ACCESSION	CQ724907.1	GI:42285764				
VERSION						
KEYWORDS						
SOURCE	Homo sapiens	(human)				
ORGANISM	Homo sapiens					
REFERENCE	1					
AUTHORS	Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.					
TITLE	Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof					
JOURNAL	Patent: WO 02068579-A 10841 06-SEP-2002;					
PE Corporation (NY) (US)						
FEATURES	Location/Qualifiers					
source	1..817					
	/organism="Homo sapiens"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:9606"					
ORIGIN						
Query Match	50.3%;	Score 817;	DB 6;	Length 817;		
Best Local Similarity	100.0%;	Pred. No. 2.5e-129;				
Matches 817;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
Qy	306	TGCGGATGTTTCATCTCCAATTCAGAAATTTACTTTTCGCAGTGGCAGGCTCTGCTGAA	365			
Db	1	TGCGGATGTTTCATCTCCAATTCAGAAATTTACTTTTCGCAGTGGCAGGCTCTGCTGAA	60			
Qy	366	TAGAAATCAGTCTGTGAACCCCTCAGACGGACATTGACGGGCTCCGGAACATCTGGATTAT	425			
Db	61	TAGAAATCAGTCTGTGAACCCCTCAGACGGACATTGACGGGCTCCGGAACATCTGGATTAT	120			
Qy	426	AAAGCCCGCGGCAAGTCCCGGGGCGGACACATAGTGTGCATGGACCGTGTGGAGAGAT	485			
Db	121	AAAGCCCGCGGCAAGTCCCGGGGCGGACACATAGTGTGCATGGACCGTGTGGAGAGAT	180			
Qy	486	CTTGGAGCTGGGAGCTGGAGACACCTCTTTCCAGGGACACAACTGGGTGCTCCAGAA	545			

Db	181	CCTGGAGCTGGCAGCTGCAGACCAACCTCTTTTCCAGGGACAAAGTGGGTGTCAGAA	240
Qy	546	GTACATCGAGAGCCCGCTGCTCTGTGTGACCAAGTTCCAGATCAGACAGTGGTTCTCT	605
Db	241	GTACATCGAGAGCCCGCTGCTCTGTGTGACCAAGTTCCAGATCAGACAGTGGTTCTCT	300
Qy	606	CGTCACGGACTGGAACCCCTCGACCATCTGTGGTTCTACAAGGAGAGTTACTTGGCGTTCTC	665
Db	301	CGTCACGGACTGGAACCCCTCGACCATCTGTGGTTCTACAAGGAGAGTTACTTGGCGTTCTC	360
Qy	666	AATCAGCGCTTCTCCCTGGACAAAGCTGACAGCGGCATCCACCTGTGCAACACCCCGT	725
Db	361	AATCAGCGCTTCTCCCTGGACAAAGCTGACAGCGGCATCCACCTGTGCAACACCCCGT	420
Qy	726	CCAGAAGTACTGAAGATGATGTGGCGCGCAGCCCTGCTGCCCGCACACAAACATGTG	785
Db	421	CCAGAAGTACTGAAGATGATGTGGCGCGCAGCCCTGCTGCCCGCACACAAACATGTG	480
Qy	786	GACCAACACAGTTTCCAGAGTACCTGACGCGCCAGGGCGCTGGCGCCGTGTGGGGCAG	845
Db	481	GACCAACACAGTTTCCAGAGTACCTGACGCGCCAGGGCGCTGGCGCCGTGTGGGGCAG	540
Qy	846	CGTCATCTACCGTTCATGAAGAAGCCATCGCCACGCCATGAAGTGGTCCCGAGGACCA	905
Db	541	CGTCATCTACCGTTCATGAAGAAGCCATCGCCACGCCATGAAGTGGTCCCGAGGACCA	600
Qy	906	CGTGGAGCTCGCAAGACAGCTTTCAGCTCTACGGGGCTGACTTCCTCTTGGAGGGA	965
Db	601	CGTGGAGCTCGCAAGACAGCTTTCAGCTCTACGGGGCTGACTTCCTCTTGGAGGGA	660
Qy	966	CTTCAGGCGCTGCTGATCGAGATCAATTCAGCCGCCACCATGCCCTCCACGCGCGT	1025
Db	661	CTTCAGGCGCTGCTGATCGAGATCAATTCAGCCGCCACCATGCCCTCCACGCGCGT	720
Qy	1026	CACGGCCAGCTGTGTGCAAGTGTGAGGAGGACACCATCAAGTGGCGTGCACCGCAG	1085
Db	721	CACGGCCAGCTGTGTGCAAGTGTGAGGAGGACACCATCAAGTGGCGTGCACCGCAG	780
Qy	1086	CTGTGACATCGGCAACTTCGAGCTCTGTGGAGGCGAG	1122
Db	781	CTGTGACATCGGCAACTTCGAGCTCTGTGGAGGCGAG	817
RESULT 2	HS355C18 101270 bp DNA linear PRI 05-JUN-2003		
LOCUS	Human DNA sequence from clone RP3-355C18 on chromosome 22q13.3		
DEFINITION	Contains the KIAA0027 gene, ESTs, STSs, GSSs and seven putative CpG islands, complete sequence.		
ACCESSION	AL022327		
VERSION	AL022327.17 GI:5304851		
KEYWORDS	HTG; CpG island; KIAA0027.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 101270)		
TITLE	Cobley,V.		
JOURNAL	Direct Submission		
COMMENT	Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jun 30, 1999 this sequence version replaced gi:5262834. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi., EMBL; Swi., SWISSPROT; Tri., TREMBL; Wp., WORMPEP; Information		

on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep -----
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
RP3-355C18 is from the library RPCI-3 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2

This sequence is the entire insert of clone RP3-355C18 The true left end of clone RP5-898I4 is at 5390 in this sequence.

FEATURES
Source

1..101270
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="22"
/map="q13.3"
/clone="RP3-355C18"
/clone_lib="RPCI-3"

misc_feature

1..1038
/note="match: STS: Em:AL022483"
complement(1..242)
/note="AluJo repeat: matches 1..237 of consensus"
358..484
/note="L1M4 repeat: matches 829..959 of consensus"
485..792
/note="AluSx repeat: matches 2..305 of consensus"
795..1095
/note="AluSx repeat: matches 1..299 of consensus"
1096..1165
/note="L1M4 repeat: matches 960..1031 of consensus"
1166..1363
/note="WSTB1 repeat: matches 4..432 of consensus"
1564..1750
/note="L1M4 repeat: matches 1031..1203 of consensus"
1757..2038
/note="AluSx repeat: matches 1..282 of consensus"
complement(2054..3342)
/note="SVA repeat: matches 1..1386 of consensus"
3343..3359
/note="2.8 copies 6 mer TCTCTT 34% conserved"
3360..3417
/note="LTR39-int repeat: matches 3904..3961 of consensus"
3418..3577
/note="L1M4 repeat: matches 1203..1363 of consensus"
3578..3861
/note="AluJo repeat: matches 1..300 of consensus"
3862..4334
/note="L1M4 repeat: matches 1363..1830 of consensus"
4341..4489
/note="L1M4 repeat: matches 2065..2311 of consensus"
complement(4492..4840)
/note="MER4A1 repeat: matches 96..472 of consensus"
4990..4999
/note="2.5 copies 4 mer CCTG 20% conserved"
complement(5142..5300)
/note="AluDb repeat: matches 137..295 of consensus"

AX301197
VERSION AX301197.1 GI:17382288
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Yue,H., Tang,Y.T., Au-Young,J., Lu,D.A., Baughn,M.R., Hillman,J.L., Azimzai,Y., Lal,P., Yao,M.G., Bandman,O., Burford,N., Batra,S., Kearney,L. and Policky,J.L.
TITLE Cytoskeleton-associated proteins
JOURNAL Patent: WO 0185942-A 39 15-NOV-2001;
Incyte Genomics, Inc. (US)
FEATURES
Source Location/Qualifiers
1..2380
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 2156553CB1"
ORIGIN
Query Match 26.7%; Score 432.8; DB 6; Length 2380;
Best Local Similarity 59.1%; Pred. No. 5.8e-84;
Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;
QY 108 CAGCAGGCAAGTGTGAAAATGCTGAGGCAAGCTCAG-GGGCCTCCCGGGGCACTTG 166
DB 361 CAGGACAAAGCAGCCAGAAACAGAGAGAAAACCCAGTGTGGTGTCCCGAGATTG 420
QY 167 TGGACATCGCGTGCAAGGTGTGCCAGGCCTA CTTGGGGCAGCTGGAGCATGAGGAATCG 226
DB 421 TGGATGAAGCTCTGTGTGGCTGCGAGAGTACCTTAGCAACTTGGGCCCATGGACATCG 480
QY 227 ACAGCTCAGCAGATCCGTGGAGGACCTCACTGAGCCGAGTGGGAGGACCTGACCCAGC 286
DB 481 ACAAGGACCTGGAGGCCCCGCTGTATCTCACCCCCGAGGGCTGGTCCCTCTCCCTCCAGC 540
QY 287 AGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTAATTTCGC 346
DB 541 GCTACTACCAAGTGTCTCAGAGGGGAGAGTCACTAGGCACCTCGACACTCAGGTCACG 600
QY 347 AGTGCCAGGCTCTGCTGAATAGAAATCACGCTGTGTGAACCTCTCAGCGGACATGACGGGC 406
DB 601 GCTGTGAGGACATCTCTGCAGCAGCTGCAGCGCGTGGTACCCAGATAGACATGGAAGGGG 660
QY 407 TCCGGAACATCTGGATTATTAAGCCCGCCGCAAGTCCCGGGCGAGACATAGTTGCA 466
DB 661 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCAAAGTCCCGTGGACGAGGCATCATGTGCA 720
QY 467 TGGACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCAACCCCTCTTTCCAGGGACA 526
DB 721 TGGACCACTGGAGGAGATCTGAACTGTGAAACCGCAACCCCGTGGTGTGATCAGGACG 780
QY 527 ACAAGTGGGTGTCCAGAAAGTACATCAGACGCGCTGTCTCATCTGTGTGACACCAAGTTCG 586
DB 781 GCAAGTGGGTGTGCAAGAAATATTTAGCGGGGCCCTCTCTCATCTTTGGCACCAGTTTG 840
QY 587 ACATCAGACAGTGTTCCTCTCAGCAGCTGGNACCCCTGACCATCTGTTCTACAAGG 646
DB 841 ACTCTCAGACAGTGTGTTCTCTGGTAACTGACTGGAACCCACTTACCCTGTGGTCTACCGCG 900
QY 647 AGAGTTACTTGGGGTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCC 706
DB 901 ACAGCTATATCCGCTTTTCCACGAGCCCTTCTCCCTGAAGAACCTGGACAACTCAGTGC 960
QY 707 ACTGTGCAACACCGCGTCCAGAAATGTA CTTGAAGAAATGATGTGGGCCGAGCCCCCTGC 766
DB 961 ACCTGTGCAACAACTCCATCCAGAAAGCAGCTGGAGAACTCATGCCACTCGGCATCCACTGC 1020
QY 767 TGGCCGACACACATGTGGAACAGCAGCAGGTTCCAGAGTACCTGACGCGCCAGGGCC 826
DB 1021 TTCCGCCGACAAACATGTGTCTAGCGAGAGGTTCCAGGAGGTTCCAGGCCCACTGTGAGGAGATGGGTG 1080

QY 827 GTGGCGCGCTGTGGGCGACGCTACCTACCTCCATGAAGAGGCGCATCGCCCCACGCCA 886
DB 1081 CCCCAAATGCTTTGGTCCACCATCATCGTGCCTGGCATGAAGGATGCTGTGATCCAGGCAC 1140
QY 887 TGAAGTGGCCCGCAGGACGACGCTGGAGCCTTCGCAAGACAGCTTTGAGCTCTACGGGGCTG 946
DB 1141 TTTCAGACCTTCCAGGACACCGTGCAGCTGTGCGAAGGCCAGCTTTGAGCTCTATGGCGCTG 1200
QY 947 ACTTCTGCTCTTGGGAGGACTTTCAGSCCTGCTGATCGAGATCAATTCAGAGCCACCA 1006
DB 1201 ACTTCTGCTTGGGAGGACTTTCAGSCCTGCTGATGATGATCAACGCGAGCCCCACGA 1260
QY 1007 TGCACCCGCTCCACGCGGCTCACGCGCCAGCTGTGTGCACAGGTGCGAGGAGCACCATCA 1066
DB 1261 TGGCACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGTGGCGTGCAGCTGACACCCCTGC 1320
QY 1067 AGGTGGCC-----GTGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTCTGT 1114
DB 1321 GCGTGTGTCATTGACCGGATGCTGGACCGCAACTGTGACACAGGAGCCTTTGAGCTCATCT 1380
QY 1115 GGAGCGAGCGGTGTGAGCGCCCGCATTCAGCGGGTCCGACCTCTCGTGGCGGGCG 1174
DB 1381 ATGAGCAGCTCTCTGTGGAGTGCCTCAATATGTGGGCATCCGGCTCTCTGTGAGAGGCT 1440
QY 1175 TCAGTGTGAGGAGAGCAGGAGGCGAGGTGCTGCTGCCCTCTTGCACACCTCAAGGCTCTGGCCT 1234
DB 1441 TCACCATCAAGAAGCCCATGGCGATGTGTCTATCGCGGATGGGGTCCCGCCAGCAGTCC 1500
QY 1235 CGCTGTGGACCGCAGCGCTGAAGGACGGGGCCCTTCGGCCATGCC-AGACCTTGC 1293
DB 1501 CTCTGTGACCCAGCGAGGCTCTGGGGAGGCAAGGACTTCGGGGATCCCTTACCCACAGT 1560
QY 1294 CAGGAGCCCCCATCAACAGCTCTCAGCGGAGCTTGGGACTCAAGAGAGAGGGGCTC 1353
DB 1561 CAGCTTCTAGGAAGGCACTGGGGCCAGGAGCTGGGGCAGCTGAGGAAGCAGTCTCA 1620
QY 1354 CCCCTGGCCCTGCTGGCACCTTAAGGGGGCAGCGAGCGGTGGAGCGCGCAGCCACAGCCC 1413
DB 1621 CTGCGCACCATTTGAGCCCGGAAAGGGAAGAAAGGCGAAAGGGCCACAGCCC 1680
QY 1414 ACCCGACCAAGCTGTCTGGGAAGGTGGAGCTCCCGGC 1451
DB 1681 TGGTCTGCCCAATCTCTGGGAGTGGGATGCCCCGAGC 1718

RESULT 4

BD158727 2326 bp DNA linear PAT 17-JAN-2003
LOCUS Primer for synthesizing full-length cDNA and use thereof.
DEFINITION BD158727
ACCESSION BD158727
VERSION BD158727.1 GI:27864485
KEYWORDS JP 2002191363-A/13570.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2326)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 13570 09-JUL-2002;
HELEX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/13570
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC

10, PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12M15/00,C12NS/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key

FEATURES	CDS	Location/Qualifiers	FT		CDS	Location/Qualifiers	source	ORIGIN
Query Match	26.6%	Score 431.6;	DB 6;	Length 2326;				
Best Local Similarity	63.5%	Pred. No. 9.3e-64;						
Matches	698;	Conservative 0;	Mismatches 389;	Indels 13;	Gaps 2;			
Qy	108	CAGCAGGCAAGATCGTGAATGCTGAGGCCAAAGCTCAG-GGGCTCCCGGGGAGCTTG	166					
Db	620	CAGAGACAGCAGCCCAAGAAACAGAGAAACCCAGTGTGGTGTCCCAAGTTTG	679					
Qy	167	TGGACATCGCGTGCAGAGTGTGCAGAGCTACCTGGGGCAGCTGGAGCATGAGGACATCG	226					
Db	680	TGGATGAAGCTCTGTGTGCGTGCAGGAGTACCTTAGCACTTCGCCCAATGACATCG	739					
Qy	227	ACAGTCAGCAGATGCGCGTGAGGACCTCACTGAGGCCGAGGTGGAGGACCTGACCCAGC	286					
Db	740	ACAAAGGACCTGGAGGCCCGCTGTACCTCACCCCGAGGGGTGGTCTCTTCTCTCCAGC	799					
Qy	287	AGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCNATTCAAGAAATTACTTTTCG	346					
Db	800	GCTACTACCAAGTGGTCCAGAGGGGCAAGAACTCAGGCACCTCGACACTCAGTCTCAGC	859					
Qy	347	AGTGCCAGGCTCTGCTGAATAGAAATCAGTCTGTGAAACCCCTCAGACGGAATTCACGGGC	406					
Db	860	GCTGTGAGGACATCTCGACAGCTGTCAGGCGGTGGTACCCCAATAGACATGAGAGGGG	919					
Qy	407	TCGCGAATCATCTGGATTATAAGCCCGCGGCAAGTCCCGGGGCGGAGACATAGTGTGCA	466					
Db	920	ATCGCAACATCTGATCGTGAAGCCAGGAGCCAAAGTCCCGCGGACGAGGCATCATGTGCA	979					
Qy	467	TGGACGTGTGGAGGATCCTGAGCTGGCAGCTGCAGACCAACCCCTTTTCAGGGACA	526					
Db	980	TGGACCACTGGAGGAGATGCTGAAGCTGTGTGAACCGGCAACCCCGTGTGATGAAGGACG	1039					
Qy	527	ACAAGTGGGTGGTCCAGAAGTACATCGAGAGCGCGCTGCTCATCTGTGACACCAAGTTTCG	586					
Db	1040	GCAAGTGGGTGGTCCAGAAGTATATTGACGGGCCCTCTCATCTTTGGACCAAGTTTG	1099					
Qy	587	ACATCAGACAGTGTTCCTCGTCAAGCATGGAACCCCGCTGACCACTCTGGTCTTACAAGG	646					
Db	1100	ACCTCAGACAGTGGTTCTCGTAACTGACTGGAACCCACTTACCGTGTGGTCTTACCGCG	1159					
Qy	647	AGATTACTTGGGTCTCAACTCAGCGCTTCTCCCTGGACAAAGCTGGACAGCGCCATCC	706					
Db	1160	ACAGCTATATCCGCTTTTCCACGAGAGCCCTTCTCCCTGAAGAACCTGGACAACTCAGTGC	1219					
Qy	707	ACCTGTGCAACAAAGCGCGTCCAGAAGTACCTGAAGAATGATGTGGCGCGACGCCCTTCG	766					
Db	1220	ACCTGTGCAACAACTCCATCCAGAAGACACCTTGGAGAACTCATGCGCATCGGCATCCATGC	1279					
Qy	767	TGCCCGCACAAATGTCGACACAGCACCAGGTTCCAGGAGTACTCTGAGCGCCAGGGGCC	826					
Db	1280	TTCCGCCAGACAAATGTTGTCTAGCCAGAGGTTCCAGGCCCACTTCAGGAGATGGTG	1339					
Qy	827	GTGGCGCGGTGTGGGCGAGCGTCACTTACCGGTCCATGAAGAAGGCCATTCGCCACGCCA	886					
Db	1340	CCCCAAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGGATGCTGTGATCCACGCAC	1399					
Qy	887	TGAAGTGGCCCGAGCACAGTGGAGCCTCGCAAGAACAGCTTTTGGTCTTACGGGGCTG	946					
Db	1400	TTACAGACTCCAGGACACCGGTGAGTGTGGAAGGGCCAGCTTTTGGCTCTATGGCGCTG	1459					
Qy	947	ACTTTCGCTTGGGAGGACCTTCAGGCCCTGGCTGTGTCAGATCAATTTTCAGGCCCAACCA	1006					

Db 680 TGGATGAAGCTCTGTGTGCGTGCAGAGGATGACCTTAGCAACCTTTGGCCCATGAGCATCG 739
QY 227 ACACGTCAGCAGATGCGGTGGAGGACCTCACTGAGCGCGAGTGGAGGACCTTGACCCAGC 286
Db 740 ACAAAGACCTGGAGGCCCGCTGTACTTACCCCCGAGGGCTGCTCTTCTCCAGC 799
QY 287 AGTACTACTCCCTCGTTCATGGCGATGCTTTCATCTCCAAATTCAGAAATTAATTTTCG 346
Db 800 GCTACTACCAAGTGGTCCAGGAAGGGGCAGAACTCAGGCACCTCCGACACTCAGGTCAGC 859
QY 347 AGTGCCAGGCTCTGCTGAATAGAAATCACTCTGTGAACCTTCAGACGGACATTTGACGGC 406
Db 860 GCTGTGAGACATCTCTGCGAGAGCTCGACGCCGTGTATCCACAGATAGACATGGAAGGG 919
QY 407 TCCGGAACATCTCGATTATAAAGCCCGCGCCCAAGTCCCGGGGCCGAGACATAGTGTGCA 466
Db 920 ATCGCAACATCTGGATCGTGAAGCCAGGAGCCAGTCCCGCGGACGAGGCATCATGTGCA 979
QY 467 TGGACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCACTCTTTCCAGGGACA 526
Db 980 TGGACCACTGGAGGAGATGCTGAAGCTGCTGAACCGCAACCCCGTGTGATGAAGGACG 1039
QY 527 ACAAGTGGGTGCTCCAGAGTACATCGACGCGCTGCTCATCTGTGACACCAATTTG 586
Db 1040 GCAAGTGGGTGCTGCAAGATATATTGAGCGGCCCTCTCTCATCTTTGGCACAAGTTG 1099
QY 587 ACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTGACCATCTGGTTCACAAGG 646
Db 1100 ACTTCAGACAGTGGTTCCTGTTAACTGATCGNACCCACTTACCGTGTGGTCTACCGCG 1159
QY 647 AGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCATCC 706
Db 1160 ACAGCTATATCCGCTTTTCCAGCGCAGCCCTTCTCCCTGAAGAACCTTGGACAACTCAGTGC 1219
QY 707 ACTGTGCAACACGCGCTCGAGAGTACTGAGAAATGATGTGGCGCCGACGCCCTCTG 766
Db 1220 ACTGTGCAACAACTCCATCCAGAAACACTGAGAACTCATGCGCATCGGATTCACACTGC 1279
QY 767 TGCCTCCGACACAACATGTGGACCAAGCAGCAGGTTCCAGAGTACCTGACGCGCCAGGGCC 826
Db 1280 TTCCGCCAGACAACATGTGCTTAGCCAGAGGTTCCAGGCCCACTGACGAGATGGGTG 1339
QY 827 GTGGGCGCTGTGGGCGACGGTCACTTACCGTTCATGAAGAAGGCCATTCGCCACGCCA 886
Db 1340 CCCCAAATGCTTGGTCCACCATCATCTGCTGCGCATGAAGGATGCTGTGATCCACGCCAC 1399
QY 887 TGAAGTGGCCCGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGCTG 946
Db 1400 TTCAGACCTCCGAGGACACCGTGCAGTGTCCGAAGCCAGCTTTGAGCTCTATGGCGCTG 1459
QY 947 ACTTCGTCCTTGGGAGGACTTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCCACCA 1006
Db 1460 ACTTCGTTTCGGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGCCCCACGA 1519
QY 1007 TGAACCGCTCCAGCGCGTCAAGGCCCAAGCTGTGTGCAAGGTGCAAGGAGGACCACTCA 1066
Db 1520 TGGCACCTCCACAGCAGTCACTGCGCGGCTCTGTGCTGGCGTGCAGAGCTGACACCTCGC 1579
QY 1067 AGGTGGCC-----GTGGACCGCAGCTGTGACATCGCACTTCGAGCTCTCTGT 1114
Db 1580 GCGTGGTCAATTGACCGGATGCTGGAGCGGCACTGTGACACAGGAGCTTTTGAGCTCATCT 1639
QY 1115 GGAGGCGAGCGGTGGTTGAGCGCGCCCACTTCAGCGGGTCCGACCTCTCGGTGGCGGGCG 1174
Db 1640 ATGACGAGCTCTGTGGAGGTGCGCTCATATATGTGGCATCGGCTCTCTGTAGAGGGCT 1699
QY 1175 TCAGTGTGAGGAGACGAG 1194
Db 1700 TCACCATCAAGAAGCCCATG 1719

RESULT 6
AK023960

LOCUS AK023960 2326 bp mRNA linear PRI 30-JAN-2004
DEFINITION Homo sapiens cDNA FLJ13898 fis, clone THYRO1001738, weakly similar to TUBULIN--TYROSINE LIGASE (EC 6.3.2.25).
ACCESSION AK023960
VERSION AK023960.1 GI:10436083
KEYWORDS Oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheroa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H., Sekine,M., Oyabashi,M., Nishi,T., Shibahara,T., Tanaka,T., Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y., Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M., Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Puruya,T., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H., Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M., Hiraoaka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S., Yosida,M., Hikota,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Oshima,A., Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Ichihara,T., Sasaki,N., Aotsuka,S., Yoshioka,Y., Matsunawa,H., Sato,N., Takami,S., Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S., Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H., Goto,Y., Shimizu,P., Wakebe,H., Hishigaki,H., Watanabe,T., Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukushima,Y., Fujimori,Y., Komiyama,M., Tashiro,H., Taniuchi,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T., Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J., Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K., Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R., Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S. Complete sequencing and characterization of 21,243 full-length human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039
2
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A. NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2326)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
Location/Qualifiers
1..2326
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYRO1001738"
/tissue_type="thyroid gland"
/clone_lib="THYRO1"


```
/note="cloning vector: pMB18SFL3"
94. >2326
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14741.1"
/db_xref="GI:10436084"
/translations="MRLNRNKKIYVERAVKQKIFTIQGCVPIRCLLRGGVWEKKM
VHSGFTLLPPOKDLSSAGDSDDTTDEDEDEBFQPSOLFDFDLDLKFDDLDGTH
ALMVGILNLRNLPWFDEVDANSFFPRCYCLGAEDEBKAFIEDFWLTAARNVLKJVK
SEWKSFTPIQAVESEAGDKPKKERNPVLVSPEFVDEALCAACEEYSLNLAHMDIDDK
LEAPLYLTPEGNSLFTQRYQVHVEGAELRHLDTQVRCEDILIQQLQAVVPOIDMGE
RNTWIVKPKASRGRIKMDHLEMLKLVNPNVPMKDKWVQKVIERPILIFGTG
FDLRWFLVTDWNPFLVFWVRSYIRFSTOPLSKNLDNSVHLNNSIOKHLNENSHR
HPLLDPNNMWSFOFALQEMGAPNAMSITIIVPGMKDAVIHALOTSOQTVOCRKASF
ELYGADFVGEDPQPMLEINASPWAPSTAVTARLACGQVADTLVVIDRMLDRNCD
THAFELIYKQPAVEPOYVIGRILLVVEGFTIKKPMWCHRRMVRPAVPLLTQKSGEA
RHFFPSLHTKAQVSPHVLHQRQVLRHQLKSLGTAKLSTLTFTAKVLTFTAKVLSL
PNNLDPKVAPSILKPRKAPALLCLRGPLEVPCCLPKSEQFLAPVGRSRPKANRSP
DCDKPRAEACPMKRLSPLKPLPLVGTGFORRRGLGDMKLGKPLLRFPFLVLVDPTPNK
KQVK"
ORIGIN
Query Match      26.6%; Score 431.6; DB 9; Length 2326;
Best Local Similarity 63.5%; Pred. No. 9.3e-64;
Matches 698; Conservative 0; Mismatches 389; Indels 13; Gaps 2;
QY 108 CAGCAGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCTCCCGGGGCGAGCTTG 166
DB 620 CAGGAGACACAGAGCCCAAGAAAACAGAGAAAACCCAGTGTGGTGTCCCCAGAGGTTTG 679
QY 167 TGCACATCGCGTGCAGAGTGTGCAGCGCTTACCTGGGGCAGCTGGAGCATGAGCAGATCG 226
DB 680 TGGATGAGCTCTGTGTGCTGCGAGAGTACCTTTAGCAACTTGGCCACATGGACATCG 739
QY 227 ACACGTCAGCAGATGCGGTGAGGACCTCACTGAGGCGGAGTGGAGGACCTGACCAGC 286
DB 740 ACAAGGACCTGGAGGCGCGCTGTACTCTACCCCGGAGGCGTGTCTCTCTCCAGC 799
QY 287 AGTACTACTCCCTGTTGATGGGATGCTTTTCACTTCAATTCAGAAATTAATCTTTGCG 346
DB 800 GCTACTACCAAGTGTGTCAGAGGGGCGAGAACTCAGGCACTTCGACACTCAGGTCAGC 859
QY 347 AGTGCCAGGCTGCTGGAATAGAAATCAGCTCTGTGAACCTCAGACGCGACATTCAGCGGC 406
DB 860 GCTGTGAGGACATCTCTGACAGCTGTCAGGCGCGTGTGTACCCCAAGATAGACATGGAAGGG 919
QY 407 TCCGGAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGCGGAGACATAGTGTGCA 466
DB 920 ATCGCAACATCTGGATCGTGAACCCAGAGCCAAAGTCCCGCGGACGAGCATCATGTGCA 979
QY 467 TGGACCGTGTGAGGAGATCTCTGAGAGTGTGCACTGACAGACCAACCTCTTTTCAGGGACA 526
DB 980 TGGACCACTTGGAGGAGATGCTGAAGTGTGTGAACGCAACCCCGTGTGATGAAGGACG 1039
QY 527 ACAAGTGGGTGTCAGAGTATACATGAGACGCGCTGTCTCATCTGTGACCAACATCTG 586
DB 1040 GCAAGTGGGTGTCAGAGATATATTAGCGCGCCCTCTCTCATCTTTGGCACCAGATTG 1099
QY 587 ACATCAGACAGTGGTTCTCTCGTCAAGGACCTGGAACCCCTGACCATCTGTTCTTCAAGG 646
DB 1100 ACCTCAGACAGTGGTTCTCTGTAACCTGACATGGAACCCCACTTACCGTGTGTTCTACCGG 1159
QY 647 AGAGTTACTTGGGTTCTCAACTCAGCGGTTCTCTCTGAGCAAGAGTGTGACAGCGCCATCC 706
DB 1160 ACAGCTATATCCGCTTTTCACGACGCGCTTCTCTCTGAGAACTTGGACAACTCAGTGC 1219
QY 707 ACCTGTGCAACACGCGCTCCAGACTACTGAGAAATGATGTGGGCGCGCAGCCCTCTG 766
DB 1220 ACCTGTGCAACAACTCCATCCAGAAAGACCTGTGAGAACTCATGCCATCGGCATCTCTG 1279
QY 767 TGCCCCCACAACAATGTGGACCAACCAAGTGTTCAGGAGTACCTGTGACGCGCCAGCGGC 826
DB 1280 TTCCGCGCAACAACATGTGGTCTAGCAGAGGTTCCAGGCCCACTTCGAGGAGATGGGTG 1339
```

```
QY 827 GTGGCGCGGTGTGGGCGAGCGTCTATCTACCGTCCATGAGAGGCGCATCGCCACGCCA 886
DB 1340 CCCAAATGCTTGGTCCACCATCATGTCCTGGCATGAAGGATGCTGTGATCCACGCAC 1399
QY 887 TGAAGGTGGGCCAGGACCAAGCTCGAGAGCTCGCAAGAACAGCTTTTGAGCTCTACGGGGCTG 946
DB 1400 TTCAGACCTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTCTATGGCGCTG 1459
QY 947 ACTTCGTCTCTTGGGAGGGAGCTTCAGGCGCTTGTGATCGAGATCAATTCAGGCCCCACCA 1006
DB 1460 ACTTCGTGTTCGGGAGGAGCTTCAGGCGCTTGTGATCGAGATCAAGCCGAGCCCCACGA 1519
QY 1007 TGCACCGTCCAGCGCGCTACGCGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCA 1066
DB 1520 TGGCACCTCCACAGCAGTCACTGCGCGGCTCTGTGCTGGCGTGCAGAGCTGACACCTTC 1579
QY 1067 AGGTGAGCC-----GTGGACCGGAGCTGTGCATCGGCAACTGTGCACAGGAGCCTTTGAGCTCATCT 1114
DB 1580 GCGTGTGTCATTGACCGGATGCTGGACCGCACTGTGCACAGGAGCCTTTGAGCTCATCT 1639
QY 1115 GGAGGCGAGCGGTGTGAGCGCCGCCCAATTCAGCGGTCGACCTCTGCGTGGCGGGCG 1174
DB 1640 ATAAGCAGCGCTGCTGTGAGGTCCTCAATATGTGGGCATCCGGCTCCTGCTAGAGGGCT 1699
QY 1175 TCAGTGTGAGGAGAGCCAGG 1194
DB 1700 TCACCATCAAGAGGCCCATG 1719
RESULT 7
HSM805098      Homo sapiens mRNA; cDNA DKFp586B0320 (from clone DKFp586B0320).
LOCUS          HSM805098      4238 bp      mRNA      linear      PRI 12-JUL-2002
DEFINITION     Homo sapiens mRNA; cDNA DKFp586B0320 (from clone DKFp586B0320).
ACCESSION      AL833939
VERSION        AL833939.1      GI:21739506
KEYWORDS       Homo sapiens (human)
SOURCE         Homo sapiens
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 4238)
AUTHORS       Wambutt R., Heubner D., Mewes H.W., Gassenhuber J. and Wiemann S.
TITLES        Direct Submission
JOURNAL       Submitted (10-JUL-2002) GSF, Institut fuer Bioinformatik MIPS,
              Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany
COMMENT       Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
              Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
              sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
              consortium of the German Genome Project.
              This clone (DKFp586B0320) is available at the RZPD in Berlin.
              Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
              Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
              information about the clone and the sequencing project is available
              at http://mips.gsf.de/proj/cDNA/.
FEATURES       Location/Qualifiers
               1..4238
                .organism="Homo sapiens"
                .mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="DKFp586B0320"
                /tissue_type="uterus"
                /clone_lib="586 (synonym: hutel). Vector pSport1; host
                DH10B; sites NotI + SalI/MluI"
                /dev_stage="adult"
               1..4238
                .gene="DKFp586B0320"
                .CDS
                .gene="DKFp586B0320"
                /note="similarity to tubulin-tyrosine ligase"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD38794.1"
```



```
/db_xref="GI:21739507"
/db_xref="GO:Q8NDN8"
/translation="MDIDKDLAPLVLPFGHSLFLQRYQVYVHGAELRHLDTOVOR
CEDILQQLQAVVPOIDMEGRNIWIKPAKSRGRGIMCNDHLEMLKLVNGPVVMK
DGRWVQKIERPLLIIFGTFDLRQWFLVDFYRDSYIRFSTQPSFLKNLD
NSVHLNCSIQRHLENSCHRHPLPPDNMWSORFOAHLQEMGAPNASTIIVPGMKD
AVIHALQTSQDTVQCRKASFLYGADVFEGDFQWMLIEINASTPMASTAVTARICA
GVOADLRVVIDRRLDRNCDTGAFELIYKQAVEVPOVYVIRLLVEGFTIKKPMAMCH
RMGVVPAPVLLTQBGSGRHHFPLSHTKAOLPSPHVLRHOGQVLRHROHSLKLVGTHA
LSTTGKALRILFTAKVIFISLPNLDLFKVAIPSLKPRKVGDL"
4216
polyA_site
/genes="DKFZp586B0320"

ORIGIN
Query Match          26.6%; Score 431.6; DB 9; Length 4238;
Best Local Similarity 63.5%; Pred. No. 8.4e-64;
Matches 698; Conservative 0; Mismatches 389; Indels 13; Gaps 2;

QY 108 CAGCAGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCCTCCGGGGCAGCTTG 166
D 1430 CAGGAGCAAGCAGCCCAAGAACAGAGAAAACCCAGTGTGGTGTCCCCAGATTG 1489
QY 167 TGGACATCGCGTCAAGGTGTGTCAGGCCTACTCTGGGGCAGCTGGAGCATGAGGACATCG 226
D 1490 TGGATGAAGCTCTGTGTGCGTGGCAGGAGTACCTTAGCAACTTGGGCCCATGGACATCG 1549
QY 227 ACACGTCAGCAGATCGGTGGAGGACCTCACTGAGCCGAGTGGGAGGACCTGACCCAGC 286
D 1550 ACAAGGACCTGGAGGGCCCGCTGTACTCACCCCCGAGGGCTGGTCCCTCTTCCTCCAGC 1609
QY 287 AGTACTACTCCCTCGTTTCATGCGGATGCTTTTCATCTCCAAATTCAGAAAATTAATTTCGC 346
D 1610 GCTACTACCAAGTGTTCAGAGGGGCGAGAACTCAGGCACCTCGACACTCAGTCCAGC 1669
QY 347 AGTGCCAGGCTGTGTGATAGAAATCACGTGTGTGAACCCCTCAGACGACATGACGGGC 406
D 1670 GCTGTGAGGACATCTCTGACGAGCTGACGCGCGTGTGATCCCGACATAGACATGGAAGGG 1729
QY 407 TCCGGAACATCTGGATTATTAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGCA 466
D 1730 ATCGCAACATCTGATCGTGAAGCCAGGAGCCAAAGTCCCGTGGACGAGGACATGATGCA 1789
QY 467 TGGACCGTGTGGAGGAGATCTCTGAGCTGGCAGCTGCAGACACACCCCTCTTTCCAGGGACA 526
D 1790 TGGACCACTGGAGGAGATCTGAAGCTGGTGAACGGCAACCCCGTGGTATCAAGAGAG 1849
QY 527 ACAAGTGGGTGTCCAGAAATGATCATGACGCGCGTGTCTCATCTGTGTGACCAAGATTGG 586
D 1850 GCAAGTGGGTGTGCAGAAATGATTAATGAGCGCGCCCTCTCTCATCTTTGGCACCAGTTTG 1909
QY 587 ACATCAGACAGTGGTTCCTCGTCACGAGCTGGNAACCCCTGACCATCTGTTCTACAAG 646
D 1910 ACCTCAGACAGTGGTTCCTCGTGAACCTGAGTGGAAACCCACTTACCGTGTGTTCTACCGCG 1969
QY 647 AGAGTTACTTGGCGTTCTCAACTCAGCGTTCTCTCCCTGGACAAGCTGGACAGCGCCATCC 706
D 1970 ACAGCTATATCCGTTTTCACGCGAGCCCTTCTCCCTGAAGAACCTGGACAACTCAGTGC 2029
QY 707 ACTGTGCAACACGCGCTCCAGAAATGACCTGAAGAAATGATGTGGGCGGAGCCCGCTGC 766
D 2030 ACCTGTGCAACAACTCCATCCAGAAAGCACTGGAGAACTCATGCCATCGGCATCCACTGC 2089
QY 767 TGGCCGCGACACACATGTGACACAGCAGCAGGTTCAGGAGTACCTGACGCGCCAGGGCC 826
D 2090 TTCCGCGACACACATGTGTCTTAGCTGAGAGGTTCAGGCGCCCACTTGCAGGAGATGGGTG 2149
QY 827 GTGGCGCCGTGTGGGCGAGCGTTCATCTAGCCGTCCTATGAAGAGGCGCATCGCCAGCCA 886
D 2150 CCCCRAATGCTTGGTCCACCATCATGTGCTGGCATGAAGATGTGTGATCCAGCAGC 2209
QY 887 TGAAGGTGGCCAGGACCAAGTGGAGCGCTCGCAAGAAACAGCTTTGAGCTCTACGGGGCTG 946
D 946
```

```
Db 2210 TTGAGACCTCCAGGACACCGGTGCGAGTGTGGAAGCCAGCTTTGAGCTCTATGGCGGTG 2269
QY 947 ACTTCTGCTCTTGGAGGGAGCTTTGAGCCCTGCTGATCGAGATCAATTTCCAGGCCCCACCA 1006
D 2270 ACTTCTGCTTGGAGGGAGCTTTCCAGCCCTGCTGATTTGAGATCAACGCCAGGCCCCACGA 2329
QY 1007 TGCACCCCGTCCAAGCGGTGTGACGGCCAGCTGTGTGCACAGGTGACAGGAGGACCATCA 1066
D 2330 TGGCACCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGCAGAGCTGACACCCCTGC 2389
QY 1067 AGTGGGCC-----GTGACCGCAGCTGTGACATCGGCACTTCGAGCTCTCTGT 1114
D 2390 GGTGTGTCATTGACCGGAGGCTGGACCGCACTGTGTGACACAGGAGCTTTGAGCTCATCT 2449
QY 1115 GGAGGCGAGCGGTGTTGAGCGCGCCCATTCAGCGGTCCGACCTCTGCTGGCGGGCGG 1174
D 2450 ATGAGCAGCTCTGCTGTGAGGTGCTTCAATATGTGGGCATCCGGCTCTCTGAGAGGGCT 2509
QY 1175 TCAGTGTGTGAGGAGAGCCAGG 1194
D 2510 TCACCATCAAGAAGCCCATG 2529

RESULT 8
HSM800637      2553 bp      mRNA      linear      PRI 18-FEB-2000
LOCUS          Homo sapiens mRNA; cDNA DKFZp434B103 (from clone DKFZp434B103);
DEFINITION     partial cds.
ACCESSION      AL096725
VERSION        AL096725.1 GI:5419858
KEYWORDS       Homo sapiens (human)
SOURCE         Homo sapiens
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 2553)
AUTHORS        Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE          Submitted (16-JUN-1999) MIPS, Am Klopferspitz 18a, D-82152
JOURNAL        Martinried, GERMANY
COMMENT        Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
                Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
                sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
                consortium of the German Genome Project.
                This clone (DKFZp434B103) is available at the RZPD in Berlin.
                Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
                Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
                information about the clone and the sequencing project is available
                at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES             Location/Qualifiers
     1..2553
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="RZPD:DKFZp434B103"
         /db_xref="taxon:9606"
         /clone="DKFZp434B103"
         /tissue_type="testis"
         /clone_lib="434 (synonym: htes3). Vector pSport1; host
         DH10B; sites NotI + SalI"
         /dev_stage="adult"
     1..2553
         /gene="DKFZp434B103"
         /gene="DKFZp434B103"
         /note="similarity to tubulin-tyrosine ligase"
         /codon_start=1
         /product="hypothetical protein"
         /protein_id="CAB46375.1"
         /db_xref="GI:5419859"
         /db_xref="GOA:O9Y4R7"
         /db_xref="UniProt/Swiss-Prot:O9Y4R7"
         /translation="MDIDKDLAPLVLPFGHSLFLQRYQVYVHGAELRHLDTOVOR
         CEDILQQLQAVVPOIDMEGRNIWIKPAKSRGRGIMCNDHLEMLKLVNGPVVMK
```


Db	1703	CCCTCTCTCATCTTTGGCACCAGTTTGACCTCAGACAGTGGTTCTCGTAACTGACTGG	1762
Qy	619	AACCCCTGACCACTCTGGTTCTCAAGAGGAGTTACTTGGGTTCTCAACTCAGCGCTTC	678
Db	1763	AACCCACTTACCGTGTGGTTCTACCGGACAGCTATATCGCTTTTCCAGCGAGCCCTTC	1822
Qy	679	TCCTTGACAGTGTGACAGCGGCATTCACCTGTGCAACAGCCGCTCCAGAATGACCTG	738
Db	1823	TCCTTGAAGAACCTGACCAACTCAGTGCACCTGTGCAACAACCTCCATCCAGAAGCACCTG	1882
Qy	739	AAGAATGATGTGGCGCGCAGCCCTGCTGCCGACACACAACATGTGGACCGACACAGG	798
Db	1883	GAGAACTCATGCCATCGGCATCCACTGCTTCCGCCAGACAACATGTGGTCTAGCCAGAGG	1942
Qy	799	TTCCAGGAGTACCTGCAGCGCCAGGGCCGTGGCGCGTGTGGGAGCAGCGCTCATCTACCG	858
Db	1943	TTCCAGGCCACCTGCAGGAGATGGGTGCCCAAAATGCTTGGTCCACCATCATCGTGCCT	2002
Qy	859	TCGATGAAGAGCCATCGCCACGCCCATGAAGGTGGCCAGGACCAACGATGGAGCCTCGC	918
Db	2003	GGCATGAAGGATGCTGTGATTCACGCACCTTCAGACCTCCAGGACACCGCTGCAGTGTGG	2062
Qy	919	AAGAACAGCTTTGAGCTCTACGGGGCTGACTTCGCTCTTGGGAGGAGACTTCAGGCCCTGG	978
Db	2063	AAGCCAGCTTTGAGCTCTATGCGCTGACTTCGTGTTGGGAGGAGACTTCAGGCCCTGG	2122
Qy	979	CTGATCGAGATCAATTCAGCCCCCACCATCGACCCCGTCACGCGCGGTCAAGGAGTGGCC	1038
Db	2123	CTGATTGAGATCAACGCCAGCCCCAGATGGCACCTCCACAGCAGTCACTGCCCGGCTC	2182
Qy	1039	TGTGCAACAGGTGAGAGAGACACCATCAAGGTGGCC-----GTGACCGCGAC	1086
Db	2183	TGTGCTGGGTGCAAGCTGACACCTCGCGCTGCTCAATTGACCGGATGCTGGACCGCAAC	2242
Qy	1087	TGTGACATCGCAACTTCAGGCTCCTGTGGAGCGACCGGT	1127
Db	2243	TGTGACACAGGAGCCCTTGAGCTCATCTATAAGACGCCCGT	2283
RESULT 10			
AX834642			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
Qy	80	GGGAGGAGCGCGGAGCAGCGACCTCAGCAGCAGGCAAGATGCTGAAAATGCTGAGGCA	139
Db	1213	GGGATCGTACAGGCGCCCTCCCTATGTGAGGAGACAGCAGCCCAAGAAACAGGAA	1272

Qy	140	AGCTCAG--GGGCTCTCCCGGGGCGAGCTTGTGTGACATCTGCGTGCAGAGGTGTGCCAGGCTTAC	198
Db	1273	AACCCAGTGTGGTGTCCCCACAGATTTGTGATGAAGCTCTGTGTGCGTGCAGGAGTAC	1332
Qy	199	CTGGGGCAGCTCGGACATGAGGACATCGACAGTCCAGCAGTGCAGGATGCCGTGGAGGACCTCACT	258
Db	1333	CTTAGCAACTTGGCCCAATGGAACATCGACAAGGACCTGGAGGCCCGCTGTACTCTCACC	1392
Qy	259	GAGGCGCAGTGGAGGAGCCTGACCCAGCAGTACTACTCCCTCGTTTCATGCGGATGCTTTTC	318
Db	1393	CCGAGGGCTGTCTCTCTCCCTCCAGCGCTACTACCAAGTGTCCACGAAGGGGAGAA	1452
Qy	319	ATCTCAATTAAGAAAATTACTTTTCGAGTGCAGGCTCTGCTGAAATGAAATCAAGTCT	378
Db	1453	CTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTGCAGCAGCTGCAGGCC	1512
Qy	379	GTGAACCCCTCAGACGAGCAATTCAGCGGCTCCGGAACATCTGATTTATAAGCCCGCGCC	438
Db	1513	GTGGTACCCAGATAGACATGGAAGGGATCGCAACATCTGGATCGTGAAGCAGAGCC	1572
Qy	439	AAGTCCCGGGCGGAGACATAGTGTGCATGGAACCGTGTGGAGGAGATCTCGAGCTGGCA	498
Db	1573	AGTCCCGTGGACGAGGCAATATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGT	1632
Qy	499	GTGCAGACCACTCTTTTCAGGAGCAACAAGTGGTGGTCCAGAAATGATACAGAGAG	558
Db	1633	AAACGCAACCCCGTGTGATGAAGGACGCGCAAGTGGTGGTGCAGAAATATATTGAGCG	1692
Qy	559	CGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTCTCTGTCACGAGCTGG	618
Db	1693	CCCTCTCTCATCTTTGGCACCAAGTTTGAATCAGACAGTGGTCTCTGTTAACTGACTGG	1752
Qy	619	AACCCCTGACCATCTGTTTCTCAAGGAGAGTTACTTTCGCGTTTCTCAACTCAGCGCTTC	678
Db	1753	AACCCACTTACCGTGTGGTCTACCGGACAGTATATCCGCTTTTCCACGAGCCCTTC	1812
Qy	679	TCCCTGGACAGCTGCAGAGCCCATTCACCTGTGCAACAGCGCGTCCAGAAATGACTGT	738
Db	1813	TCCCTGAAGAACCTGGACAACTCAGTGCACCTGTGCAACAACTCCATCCAGAAACCTGT	1872
Qy	739	AGAATGATGTGGCGCGCAGCGCTGCTGCCCGCACACAACTGTGGACAGCAGCAGG	798
Db	1873	GAGAATCATGCGCATCGGCATCCACTGTTCCGCCAGACAACATGTGTGTAGCCAGAG	1932
Qy	799	TTCCAGGAGTACTGTGCAGCGCGCAGCGCGTGGCGCGCTGTGGGAGCAGCGTCACTACCG	858
Db	1933	TTCCAGGCCACCTGCAGGAGATGGTGGCCCAATGCTTGGTCCACCATCATCTGTCCT	1992
Qy	859	TCCATGAAGAGGCCATCGCCCAACGCCATGAAGTGGCGCCAGGACCAAGTGGAGCTCGC	918
Db	1993	GGCATGAAGGATGTGTGATCCACGCACCTTCAGACCTCCAGGACACCGTGCATGTCCG	2052
Qy	919	AAGAAGCTTTGAGCTCTAGCGGCTGACTTTCGCTTGGGAGGAGCTTTCAGGCGCTGG	978
Db	2053	AAGCCAGCTTTGAGCTCTATGGCGCTGACTTTCGTTTCGGGGTGGACTTTCAGGCGCTGG	2112
Qy	979	CTGATCGAGATCAATTCAGCGCCCACTGACACCGCTCCACCGCTCCAGCCCGCAGCTG	1038
Db	2113	CTGATTGAGATCAACGCCAGCCCACTGAGTGGACCCCTCCACAGCAGTCACTCCCGGCTC	2172
Qy	1039	TGTGACAGGTGTCAGGAGGACACCATCAAGTGGCGCTGGAGCCCGAGCTGTGAC	1092
Db	2173	TGTGTCGCGTGAAGCTGACACCCCTGGCGGTGGTCAATTGACCGGAGGTGGAC	2226
RESULT 11			
AX097236			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			

AK097236 2848 bp mRNA linear PRI 30-JAN-2004
Homo sapiens cDNA FLJ39917 fis, clone SPLEN2019405, highly similar to Homo sapiens FHTL protein mRNA.
AK097236
AK097236.1 GI:21756925
oligo capping; fis (full insert sequence).

SOURCE	Homo sapiens (human)	ORIGIN	Query Match	25.8%;	Score 418;	DB 9;	Length 2848;	
ORGANISM	Homo sapiens		Best Local Similarity	63.9%;	Pred. No. 1.9e-61;			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		Matches	648;	Conservative	0;	Mismatches 365; Indels 1; Gaps 1;	
AUTHORS	1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oabayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, S., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hotta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Sato, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.							
TITLE	Complete sequencing and characterization of 21,243 full-length human cDNAs							
JOURNAL	Nat. Genet. 36 (1), 40-45 (2004)							
PUBMED	14702039							
REFERENCE	2							
AUTHORS	Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.							
TITLE	NEDO human cDNA sequencing project							
JOURNAL	Unpublished							
REFERENCE	3 (bases 1 to 2848)							
AUTHORS	Isogai, T. and Yamamoto, J.							
TITLE	Direct Submission							
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.							
FEATURES	Location/Qualifiers							
source	1..2848 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="SPLEN2019405" /tissue_type="spleen" /clone_lib="SPLEN2" /notes="cloning vector: pME18SFL3"							


```
RESULT 12
AF078842
LOCUS 1684 bp mRNA linear PRI 09-JAN-2000
DEFINITION Homo sapiens HOTT1 protein mRNA, complete cds.
ACCESSION AF078842
VERSION AF078842.1 GI:6683744
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,
Zhou, S., Liu, M. and He, F.
TITLE Functional prediction of the coding sequences of 50 new genes
deduced by analysis of cDNA clones from human fetal liver
Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 1684)
AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,
Zhou, S., Liu, M. and He, F.
Direct Submission
TITLE Submitted (17-JUL-1998) Dept. of Experimental Hematology, Beijing
Institute of Radiation Medicine, 27 Taiping Rd, Beijing 100850,
P.R.China
FEATURES
source
1..1684
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HQ0207"
/tissue_type="liver"
/dev_stage="fetus"
1..398
399..1277
/notes="hnp0207; similar to bovine and pig tubulin-tyrosine
ligase (TTL): Swiss-Prot Accession Numbers P38584 and
P38160"
/codon_start=1
/evidence=not experimental
/product="HOTT1 protein"
/protein_id="AAF23353.1"
/db_xref="GI:6683745"
/translation="MEGDRNFVFKPAKSRGRCIMDHLEMLKLVNGNPVVMKDG
KWVQYIERPLLIQTKFDLQWFLVTDNPLTVFYRDSVIRFSTQPSKLNLDNS
VHLNNSIQKHLENSCHREPLLPDNNWSSQRFQAHLOEMGAPNAWSTIIVRKMADV
IHALQTSQTVQCRKASFELYGADFVFGEDFQFLIEINASTPNAFTAVTARLCAGV
QADTLRVVDRMDLRNCDTGAFELIYKQPVTTSPASTPRPSCLLPMYSIDTRASSDDS
TASWALRPFQAPR"
1278..1684
3'UTR
ORIGIN
Query Match 23.4%; Score 380.2; DB 9; Length 1684;
Best Local Similarity 68.6%; Pred. No. 5.6e-55;
Matches 545; Conservative 0; Mismatches 238; Indels 12; Gaps 1;
QY 345 GCAGTCCAGGCTCTGCTGAATAGATCAGCTGTGTGAACCTTCACAGCGACATTGACGG 404
Db 347 GCGCTGTGAGGACATCTTCAGCAGCTGCGAGCGCGTGTACCCAGATAGATGAAGG 406
QY 405 GCTCCGGAACATCTGATTATTAAGCCGCGGCGCAAGTCCCGGGCGGAGACATAGTGTG 464
Db 407 GGATCGCAACTTCTGATCTGTGAAGCCAGAGCCAAAGTCCCGCGGACGAGGCATCATGTG 466
QY 465 CATGGACCGTGTGGAGGAGATCTCGAGCTGGCAGCTGCAGACCACCTCTTTCCAGGGA 524
Db 467 CATGGACCACTTGGAGGAGATGCTGAAGCTGGTGAACCGCAACCCCGTGTGATGAAGA 526
QY 525 CAACAAGTGGGTGGTCCAGAAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGTT 584
Db 527 CGGCAAGTGGGTGGTCCAGAGTATATTGAGCGGCGCCCTCTCTCATCTTTGGACCAAGTT 586
```

TITLE

Generation and initial analysis of more than 15,000 full-length

```
QY 585 CGACATCAGACAGTGTTCTCTCGTCACGAGCTGGAAACCCCTGACCATCTGGTTCTACAA 644
Db 587 TGACCTCTCAGACAGTGTTCTCTCGTAACTGACTGGAACCCACTTACCGTGGTTCTACCG 646
QY 645 GGAGAGTTACTTTCGGTGTCTCAACTCAGCGCTTCTCCCTGGACAAGCTCGACAGCGCAT 704
Db 647 CGACAGCTATATCCGCTTTTCACGAGCGCCCTTCTCCCTGAAAGAACCCTGGACAACCTCAGT 706
QY 705 CCACCTGTGCAACAACCGCGTCCAGAGTACCTCGAAGAATGATGTGGCGCGCAGCGCCCT 764
Db 707 GCACCTGTGCAACAACCTCCATCCAGAGCACTCGAGAACTCATGCCATCGGCATCCACT 766
QY 765 GTGCCCGGCACACAATGTGACACGACACAGGTTCCAGGAGTACCTCGACGCGCAGGG 824
Db 767 GCTTCCGCGCAGACAACATGTGGTGTAGCCAGAGGTTCCAGGCGCCACCTGACAGAGATGG 826
QY 825 CCGTGGCGCGTGTGGGCGAGCGTCTATCTACCGTCCATGACCGTCCATGAAGAAGCCATCGCCACGC 884
Db 827 TGCCCAAAATGCTTGGTCCACCATCATGTGCTGGCATGAAGGATGCTGTGATCCACGC 886
QY 885 CATGAAGTGTGGCCAGGACCACTGGAGCTCGCAAGAACACAGCTTTAGCTCTACGGGCG 944
Db 887 ACTTCAGACCTCCAGAGCACCGTGCAGTGTGGAAGCCAGCTTTAGCTCTATGGCGC 946
QY 945 TGACTTGCTGTGGAGGAGACTTCAGGCCCTGGCTGTGATCGAGATCAATTCAGGCCCCAC 1004
Db 947 TGACTTGTGTTCGGGAGGAGACTTCAGGCCCTGGCTGTGATGAGATCAACGCCAGCCCCAC 1006
QY 1005 CATGACCGCTCCAGCGCGTCCAGCGCCAGCTGTGTGCACAGGTGCGAGGAGGACACCAT 1064
Db 1007 GATGGACCCCTCCAGCAGCTCACTGCGCGGCTCTGTGTGGCTGCAAGCTGTGACACCT 1066
QY 1065 CAAGGTGGCC-----GTGGAACGCAAGCTGTGACATCGGCAACTTCGAGTCTCT 1112
Db 1067 GCGCGTGGTCATTGACCGGATGCTGGAACGCAACTGTGACACAGGAGCTTTGAGTCTCAT 1126
QY 1113 GTGAGGCGAGCGCGT 1127
Db 1127 CTATAAGCAGCGCGT 1141
```

```
RESULT 13
BC006830
LOCUS
DEFINITION Mus musculus 1897 bp mRNA linear ROD 03-OCT-2003
IMAGE:3597662, complete cds.
ACCESSION BC006830
VERSION BC006830.1 GI:13905089
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1897)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahney,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalek,U., Smalley,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
```


human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
 2388257
 12477932
 2 (bases 1 to 1897)
 Direct Submission
 Strausberg, R.
 Submitted (27-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: csapbe-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 16 Row: g Column: 10.
 Location/Qualifiers
 1. 1897
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:11856 IMAGE:3597662"
 /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
 ductal carcinoma. 5 month old virgin mouse."
 /clone_lib="NCI CGAP Mam6"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1. 1897
 /gene="4833441J24Rik"
 /db_xref="LocusID:101100"
 /db_xref="MGI:2141418"
 590. 1390
 /gene="4833441J24Rik"
 /codon_start=1
 /product="4833441J24Rik protein"
 /protein_id="AAH06830.1"
 /db_xref="GI:13905090"
 /db_xref="LocusID:101100"
 /db_xref="MGI:2141418"
 /translation="MEGDNIWIKPKAKSRGIMCMNPLDEMLKLVDCNPMLKDG
 KWIQYIERPLIFGKFLRQWFLVDNPLTVFYRDSYRFTQPSLKNLDS
 VHLNNSIQHLSRCHRHPLPDNWSRFOALHQVEDAPKWSVIVPMKAIV
 IHALQTSQDNVRCKAFELYGDFVEDFPQLIEINASPMPASTAVRLCAGV
 QADTLRVVIDRRCSDTGAFLIYKQVRCSTSHKQPHGPGIWCTE"
 590. 1270
 /gene="4833441J24Rik"
 /note="TTL; Region: Tubulin-tyrosine ligase family.
 Tubulins and microtubules are subjected to several
 post-translational modifications of which the reversible
 deetyrosination/tyrosination of the carboxy-terminal end of
 most alpha-tubulins has been extensively analysed. This
 modification cycle involves a specific carboxypeptidase
 and the activity of the tubulin-tyrosine ligase (TTL). The
 true physiological function of TTL has so far not been
 established. Tubulin-tyrosine ligase (TTL) catalyses the
 ATP-dependent post-translational addition of a tyrosine to
 the carboxy terminal end of deetyrosinated alpha-tubulin.
 In normally cycling cells, the tyrosinated form of tubulin
 predominates. However, in breast cancer cells, the

deetyrosinated form frequently predominates, with a
 correlation to tumour aggressiveness. On the other hand,
 3-nitrotyrosine has been shown to be incorporated, by TTL,
 into the carboxy terminal end of deetyrosinated
 alpha-tubulin. This reaction is not reversible by the
 carboxypeptidase enzyme. Cells cultured in 3-nitrotyrosine
 rich medium showed evidence of altered microtubule
 structure and function, including altered cell morphology,
 epithelial barrier dysfunction, and apoptosis"
 /db_xref="CDD:pfam03133"
 ORIGIN
 Query Match 22.6%; Score 366.4; DB 10; Length 1897;
 Best Local Similarity 59.3%; Pred. No. 1.2e-52;
 Matches 499; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
 QY 379 GTGAACCTCTCAGACGACATTTGACGGGCTCCGGAACATCTGGATTATAAAGCCGCGGCC 438
 DB 572 GTGGTACCCAGTTAGACATGAGGGGGATCAGAACATCTGGATCGTGAAGCCCGGAGCC 631
 QY 439 AAGTCCGGGGCCGACACATAGTGTGATGGACCGTGTGGAGGAGATCTTGGAGCTGGCA 498
 DB 632 AAGTCCGGGGCCGAGGGGATTTATGTGCATGAACCGCTTGGATGAGATGCTGAAGCTGGT 691
 QY 499 GCTGCAGACCACTCTTCCAGGGACAACAAGTGGTGTCCAGAGTACATCGAGACG 558
 DB 692 GACTGCAACCCCATGCTCATGAAGATGGCAAGTGGATCGTCAGAGATACATTTGACGG 751
 QY 559 CCGCTGCTATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCACGGACTGG 618
 DB 752 CCCTGCTTATCTTCGGCCACCAAGTTTGACCTGAGACAGTGGTTCCTGGTGACTGCTG 811
 QY 619 AACCCCTGACCATCTGGTTCTACAAGAGAGTTACTTTGGCGTTCTCAACTCAGCGGTTTC 678
 DB 812 AACCCACTACCGTGTGGTTTACCGAGACAGCTACATTCGTTCTTCCACACAGCCCTTC 871
 QY 679 TCCTTGGAAGCTGACAGCGGCATCCACCTGTGCAACAACCGCCGTCCAGAGAGTACTCTG 738
 DB 872 TCCTTGAAGAACCTGGACAACTCTGTGCACCTGTGTAACAACCTCCATCCAGAGACACTT 931
 QY 739 AAGAATGATGTGGGCGCAGCCCTCTGCTGCCCGCACACAACATGTGGACAGCAGCCAGG 798
 DB 932 GAGGCTTCCTGTACCGGCACCGATGCTGCCCCAGATPAACATGTGGTCCAGCCAGAGG 991
 QY 799 TTCAGAGGTACTCTGACGCGCCAGGCGCGTGGGCGCGTGTGGGGCAGCGTCTATACCCG 858
 DB 992 TTTTCAGGCCCATTTGCAGGAGGTAGATGCCCCAAAGGCCCTGTGTCAGCGTCTATCGTCCA 1051
 QY 859 TCCNTGAAGAGCCCATCGCCACGCCATGAAGTGGCCGAGCAGCAGCCAGTGGAGCCCTCG 918
 DB 1052 GGCATGAAGGCTGCTGTGATCCACGCCCTGTCAGACCTCCCAAGACAAACGTGCAAGTGC 1111
 QY 919 AAGAAGAGCTTTGAGCTCTACGGGGGTGACTTCTGCTTGGGAGGAGACTTCAGGCGCTCG 978
 DB 1112 AAGCCAGCTTTGAGCTCTATGGGGCAGACTTTGTGTTGGGGAAGACTTCCAGCCCTGG 1171
 QY 979 CTGATCGAGATCAATTCAGCCCCACCATGCACCCCGTCCACGCGGTTCACGGGCCAGCTG 1038
 DB 1172 TTGATTGAATCAATGCCAGCCCCACCATGGCACCTTCACAGGCTGTCTACTGCCCGCTC 1231
 QY 1039 TGTGCAGAGTGCAGAGACACCATCAAGGTGGCGTGGACCGCAGCTGTGACATCGCG 1098
 DB 1232 TGTGCCGCTGTGCAGGACAGATACCTCTGTCGTCATCGACCGCGCTTGGACCGCAGC 1291

RESULT 14

AC119959
 LOCUS 163783 bp DNA linear ROD 23-JUL-2004
 DEFINITION Mus musculus chromosome 15, clone RP24-467H19, complete sequence.
 AC119959
 ACCESSION
 VERSION AC119959.8 GI:50540805
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 163783)

Birren B., Nusbaum, C. and Lander, E.

TITLE

Mus musculus chromosome 15, clone RP24-467H19

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 163783)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lanazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Ribback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

3 (bases 1 to 163783)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (26-MAY-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

4 (bases 1 to 163783)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (23-JUN-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 23, 2004 this sequence version replaced gi:47679219.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L25717

Center clone name: 467_H_19

----- Location/Qualifiers

1..163783

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="15"

/map="15"

/clone="RP24-467H19"

/clone_lib="RPCI-24 Male Mouse BAC"

1..4

/note="clone boundary"

clone end:SP6

site:MbolI

complement(169..204)

/rpt_family="ORR1B1"

818..994

/rpt_family="ORR1D"

1312..1527

/rpt_family="B3"

1573..1645

/rpt_family="ID3"

1665..1793

/rpt_family="B1P"

1794..1959

/rpt_family="B3"

complement(2256..2658)

/rpt_family="LI"

complement(2659..2833)

/rpt_family="B2_Mm2"

complement(2834..3302)

/rpt_family="LI"

complement(3311..3448)

/rpt_family="B1_MM"

complement(3449..4218)

/rpt_family="LI"

4214..4557

/rpt_family="Lx6"

5120..5279

/rpt_family="CT-rich"

complement(5297..5423)

/rpt_family="B1P"

complement(5451..5524)

6016..6494

/rpt_family="RLR11A"

complement(6678..6797)

/rpt_family="B1_MM"

complement(6810..6945)

/rpt_family="B1_MM"

7008..7148

repeat_region


```
repeat_region      /rpt_family="B1_MM"
                    complement(7763..7941)
repeat_region      /rpt_family="B3"
                    complement(7960..8105)
repeat_region      /rpt_family="B1_MM"
                    complement(8121..8300)
repeat_region      /rpt_family="B3A"
                    846..8625
repeat_region      /rpt_family="B4A"
                    8618..8786
repeat_region      /rpt_family="B1_MM"
                    8804..9102
repeat_region      /rpt_family="L1MB2"
                    9162..9217
repeat_region      /rpt_family="PB1D7"
                    complement(9339..9460)
repeat_region      /rpt_family="B3"
                    complement(9598..9748)
repeat_region      /rpt_family="B3"
                    complement(9755..12502)
repeat_region      /rpt_family="L1_MM"
                    12503..12992
repeat_region      /rpt_family="TAPLTR2_MM"
                    12994..13717
repeat_region      /rpt_family="TAPLTR2_MM-int"
                    13718..13836
repeat_region      /rpt_family="TAPLTR2_MM-int"
                    13832..14083
repeat_region      /rpt_family="TAPLTR2_MM-int"
                    14083..15904
repeat_region      /rpt_family="TAPLTR2_MM-int"
                    15905..16394
repeat_region      /rpt_family="TAPLTR2_MM"
                    complement(16395..20012)
repeat_region      /rpt_family="L1"
                    complement(20013..20157)
repeat_region      /rpt_family="B1_MM"
                    20192..20290
repeat_region      /rpt_family="B2_Mm2"
                    20410..20553
repeat_region      /rpt_family="B1_MM"
                    20586..20777
repeat_region      /rpt_family="B2_Mm2"
                    21012..21082
repeat_region      /rpt_family="(CAA)n"
                    21192..21294
repeat_region      /rpt_family="PB1D10"
                    21301..21342

Query Match      17.5%; Score 283.6; DB 10; Length 163783;
Best Local Similarity 76.7%; Pred. No. 7.3e-39;
Matches 365; Conservative 0; Mismatches 99; Indels 12; Gaps 1;

Qy      685  GACAAAGTGGACAGCGCCATCCACTGTGCAACACCGCGTCCAGAGAGTACTCGAAGAT 744
Db      136973  GACATTGCCACAGCGGCATCCACTGTGTAACTCCATCCAGAGCGGCTCAAAAT 137032

Qy      745  GATGTGGCGCGACGCCCTGTCTGCCCGCACACAACATGTGGACACGACGAGTTCAG 804
Db      137033  GATAAAGAGCGCAGTCCGCTGCTACCTTGTCTAACTATGAGGACGACGCTCGTTTCCAG 137092

Qy      805  GAGTACCTGACGCGCAGCGCGGTGGCGCGTGTGGGCGACGCTCATCTACCGTCCATG 864
Db      137093  GAGTACCTGCAAGAGAGGGCGGAGGAGAACTGTGGGGGTAGCATCATCTACCGTCTATG 137152

Qy      865  AAGAAGGCCATCGCCCGCCATCAAGTGTGCGCCAGGACGACGCGGAGCTTCGCAAGAAC 924
Db      137153  AAAAGAGCTGTACAAATGCTCAGGTAGTCCCGGACGACGACGAGAGCCCGTAAAGAAC 137212

Qy      925  AGCTTTGAGCTCTACGGGGCTGACTTCGTCTTGGGAGGAGATTTCAGGCCCTGCGTATC 984
Db      137213  AGCTTCGAGCTCTACGGAGCGGACTTCATCTCTGGGGCGAGACTTCAAGCCATTGCTTATC 137272
```

```
Qy      985  GAGATCAATTTCCAGCCCGCCACCATGCACCGCTCCACGCGGTCCAGCGCCAGCTGTGTGCA 1044
Db      137273  GAGATCAATTTCCAGCCCGCCACCATGCACCGCTCCACCGTCCAGCTGTGTGTC 137332

Qy      1045  CAGGTGCGAGGAGGACACCATCAAGGTGGCGCG-----TGGACCGCAGCTGTGAC 1092
Db      137333  CAGGTGCGAGGAGGACACCATCAAGGTGGTGGATCGCAAACTGGACCGCAAACTGTGAC 137392

Qy      1093  ATCGGCAACTTCGAGCTCCTGTGAGGCGACCGGTGGTTGAGCGCGCCCGCCATTCAG 1148
Db      137393  ATCGGCAACTTTGAGCTTCTGTGCGCGCAGGTGAGGATAGCAGCTACCTCTTTAG 137448

RESULT 15
AC117700      201420 bp      DNA      linear      HTG 26-SEP-2004
LOCUS      Mus musculus chromosome 15 clone RP23-455J20 map 15, *** SEQUENCING
DEFINITION      IN PROGRESS ***, 3 unordered pieces.
ACCESSION      AC117700
VERSION      AC117700.17 GI:52694698
KEYWORDS      HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 201420)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE      Mus musculus chromosome 15, clone RP23-455J20
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 201420)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
Landers,T., Lebecky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE      Direct Submission
JOURNAL      Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE      3 (bases 1 to 201420)
AUTHORS      Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B.,
DeArelano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Illiev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Minova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
```


Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-SEP-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 26, 2004 this sequence version replaced gi:52077711.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L24098
Center clone name: 455_J_20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 82149: contig of 82149 bp in length
* 82150 82249: gap of unknown length
* 82250 98003: contig of 15754 bp in length
* 98004 98103: gap of unknown length
* 98104 201420: contig of 103317 bp in length.
Location/Qualifiers
1. .201420
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-455J20"
/clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Query Match 17.5%; Score 283.6; DB 2; Length 201420;
Best Local Similarity 76.7%; Pred. No. 7.1e-39;
Matches 365; Conservative 0; Mismatches 99; Indels 12; Gaps 1;

QY 685 GACAAAGCTGGACAGCGCCATCCACCTGTGTGCAACACGCCGTCAGAAAGTACCTGAAGAAT 744
Db 4228 GACATTGCCCAACAGCGCATCCACCTGTGTGTAACAACATCCATCCAGAGGGCGCTCAAAAAT 4287

QY 745 GATGTGGCGCGACGCCCTGCTGCCCGCACACAACATGTGGACAGCACCGAGTTCCAG 804
Db 4288 GATAAAGAGCGCAGTCCGCTGTACCTTGTATACATGTGGACACGACTCGTTTCCAG 4347

QY 805 GAGTACTCTGAGCGCGCAGCGCGCTGTGGCGCAGCGTCATCTACCGTCCATG 864
Db 4348 GAGTACTCTGAGAGAGGGGCGCAGAGGAGACGTGGGTAGCATCATCTACCGCTCTATG 4407

QY 865 AAGAAGGCCATGCCCAACGCGCATGAAGGTGGCCAGGACCAACGTGGAGCCTCGCAAGAAC 924
Db 4408 AAAAGAGCTGTACCAATGCCATGAGGGTAGCCAGGACCAACGTAGAGCCCGTAAGAAC 4467

QY 925 AGCTTTGAGCTCTACGGGGCTGACTTCGTTCTGGGAGGACTTCAGGCCCTGGCTGATC 984
Db 4468 AGCTTCGAGCTCTACGGAGCGGACTTCATCCTGGGGCGAGACTTCAAGGCATGGCTTATC 4527

QY 985 GAGATCAATTCAGAGCCCGACCATGACCCCGTCACCGCGGTACCGGCCAGCTGTGTGCA 1044
Db 4528 GAGATCAACTCCAGAGCCCGACCATGACCCCGTCACCTCCCGTCACAGCTCAGCTCTGTGCC 4587

QY 1045 CAGGTGAGGAGGACACCATCAAGGTGGCGG-----TGGACCGAGCTGTGAC 1092

Db 4588 CAGGTGACAGGAGGACACCATCAAGGTGTGGATCGCAAACTGGACCGAAACTGTGAC 4647
QY 1093 ATCGGCAACTTCGAGCTCCTGTGGAGGACCGCGGTGTGAGCGCGCCCAATTTCAG 1148
Db 4648 ATCGGCAACTTTGAGCTTCTGTGGCGCAGGTGAGGATAGCAGCTACCCCTTCTTAG 4703

Search completed: September 25, 2005, 11:26:54
Job time : 7536.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 23:48:14 ; Search time 1417.93 Seconds
(without alignments)
8666.982 Million cell updates/sec

Title: US-10-615-659-1
Perfect score: 1838
Sequence: 1 tggagcagcctggggccc.....actctcccaaggcgaattc 1838

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq: *
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq: *
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq: *
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq: *
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq: *
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq: *
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq: *
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq: *
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq: *
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq: *
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1838	100.0	1838	19	US-10-615-659-1
2	1838	100.0	1838	19	US-10-615-659-1
3	1782.8	97.0	3554	19	US-10-615-659-12
4	1782.8	97.0	3554	19	US-10-615-659-12
5	1678.8	91.3	3465	19	US-10-615-659-11
6	1678.8	91.3	3465	19	US-10-615-659-11
7	1677.2	91.3	1939	19	US-10-615-659-9
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 12, Appl
					Sequence 12, Appl
					Sequence 11, Appl
					Sequence 11, Appl
					Sequence 9, Appli

8	1677.2	91.3	1939	19	US-10-615-659-9	Sequence 9, Appli
9	1574.8	85.7	1859	19	US-10-615-659-10	Sequence 10, Appl
10	1574.8	85.7	1859	19	US-10-615-659-10	Sequence 10, Appl
11	726	39.5	726	19	US-10-615-659-3	Sequence 3, Appli
12	726	39.5	726	19	US-10-615-659-3	Sequence 3, Appli
13	501	27.3	101270	20	US-10-723-860-631	Sequence 631, App
14	432.8	23.5	2380	18	US-10-275-595A-39	Sequence 39, Appl
15	427	23.2	490	16	US-10-029-386-24894	Sequence 24894, A
16	421.4	22.9	2538	15	US-10-102-524-1765	Sequence 1765, A
17	421.4	22.9	2553	21	US-10-956-157-2016	Sequence 2016, Ap
18	421.4	22.9	2553	22	US-10-756-149-1971	Sequence 1971, Ap
19	421.4	22.9	3001	15	US-10-037-270-494	Sequence 494, App
20	421.4	22.9	3001	17	US-10-117-722-494	Sequence 1765, Ap
21	418	22.7	2848	17	US-10-108-260A-1766	Sequence 1765, Ap
22	413	22.5	2611	9	US-09-925-298-199	Sequence 199, App
23	413	22.5	2611	14	US-10-102-806-199	Sequence 199, App
24	387.6	21.1	526	13	US-10-027-632-135927	Sequence 135927,
25	387.6	21.1	526	13	US-10-027-632-135928	Sequence 135928,
26	387.6	21.1	526	17	US-10-027-632-135927	Sequence 135927,
27	387.6	21.1	526	17	US-10-027-632-135928	Sequence 135928,
28	381.8	20.8	1400	21	US-10-956-157-9356	Sequence 9356, Ap
29	381.8	20.8	1673	21	US-10-956-157-4121	Sequence 4121, Ap
30	319.8	17.4	1400	21	US-10-956-157-7251	Sequence 7251, Ap
31	272.4	14.8	2241	19	US-10-615-659-27	Sequence 27, Appl
32	272.4	14.8	2241	19	US-10-615-659-27	Sequence 27, Appl
33	196	10.7	755	13	US-10-027-632-135929	Sequence 135929,
34	196	10.7	755	13	US-10-027-632-135930	Sequence 135930,
35	196	10.7	755	17	US-10-027-632-135929	Sequence 135929,
36	196	10.7	755	17	US-10-027-632-135930	Sequence 135930,
37	190.6	10.4	1958	21	US-10-956-157-4903	Sequence 4903, Ap
38	166	9.0	1728	22	US-10-450-763-5962	Sequence 5962, Ap
39	164.4	8.9	2250	22	US-10-450-763-5962	Sequence 5962, Ap
40	163.4	8.9	418	9	US-09-983-965-2032	Sequence 2032, Ap
41	131.8	7.2	492	10	US-09-918-995-35253	Sequence 35253, A
42	130	7.1	525	16	US-10-029-386-11174	Sequence 11174, A
43	122.6	6.7	436	17	US-10-242-535A-7660	Sequence 7660, Ap
44	122.6	6.7	436	18	US-10-085-783A-7660	Sequence 7660, Ap
45	113.4	6.2	279	18	US-10-424-599-119452	Sequence 119452,

ALIGNMENTS

RESULT 1
US-10-615-659-1
; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1

Query Match 100.0%; Score 1838; DB 19; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGAGCAGCCTGGGGCCCATCGCTATGAGCGCGGAAGTCGGCGGTTGGGAGCC 60
|||||

RESULT 3
US-10-61
; Sequen
; Public
; GENERA
; APPLI
; TITLE
; TITLE
; FILE
; CURRE
; CURRE
; PRIOR
; PRIOR
; NUMBE

961	AC	TG	CAG	CGCC	CAG	GGCG	CGT	TGG	CGCG	TGT	TGG	GGCG	CAG	CGT	CAT	CT	AC	CG	CGT	CAT	GN	A	GA	1021							
1021	AG	CC	AT	CG	CCCA	CG	CCAT	GA	AG	GT	GG	CC	CAG	ACCA	CG	TG	GG	AG	CG	CT	CG	CA	GA	1080							
1021	AG	CC	AT	CG	CCCA	CG	CCAT	GA	AG	GT	GG	CC	CAG	ACCA	CG	TG	GG	AG	CG	CT	CG	CA	GA	1080							
1081	TT	G	AG	CT	CT	AC	GG	GG	CT	G	ACT	CT	CG	CT	TT	T	GG	A	GG	CA	CT	T	C	GA	1140						
1081	TT	G	AG	CT	CT	AC	GG	GG	CT	G	ACT	CT	CG	CT	TT	T	GG	A	GG	CA	CT	T	C	GA	1140						
1141	TC	AA	TT	CC	AG	CCCC	CA	AT	GC	AC	CC	GT	CA	CG	CG	CG	GT	CA	CG	CG	CG	GT	CG	CA	1200						
1141	TC	AA	TT	CC	AG	CCCC	CA	AT	GC	AC	CC	GT	CA	CG	CG	CG	GT	CA	CG	CG	CG	GT	CG	CA	1200						
1201	TG	CA	G	AG	AG	CA	CC	AT	CA	AG	GT	GG	CG	TG	CA	CG	CA	GT	TG	CA	CG	CA	CT	T	GA	1260					
1201	TG	CA	G	AG	AG	CA	CC	AT	CA	AG	GT	GG	CG	TG	CA	CG	CA	GT	TG	CA	CG	CA	CT	T	GA	1260					
1261	TC	CT	GT	TG	AG	GG	CAG	CG	GT	TG	TT	T	AG	CG	CG	CC	CA	TT	T	AG	CG	GG	T	CG	AG	1320					
1261	TC	CT	GT	TG	AG	GG	CAG	CG	GT	TG	TT	T	AG	CG	CG	CC	CA	TT	T	AG	CG	GG	T	CG	AG	1320					
1321	CG	GG	CT	CA	GT	TG	AG	AG	AG	CA	CG	AG	CG	CA	GT	TG	T	CG	CC	GT	T	CG	CA	CT	CA	1380					
1321	CG	GG	CT	CA	GT	TG	AG	AG	AG	CA	CG	AG	CG	CA	GT	TG	T	CG	CC	GT	T	CG	CA	CT	CA	1380					
1381	CG	GG	CT	CG	CT	TG	TT	TG	AG	CG	CG	CA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	1440					
1381	CG	GG	CT	CG	CT	TG	TT	TG	AG	CG	CG	CA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	1440					
1441	CT	GC	CC	AG	GG	AC	CCCC	AT	CA	CC	AG	CT	CT	CA	GG	CG	CA	CT	TG	GG	AG	CT	GA	AG	AG	1500					
1441	CT	GC	CC	AG	GG	AC	CCCC	AT	CA	CC	AG	CT	CT	CA	GG	CG	CA	CT	TG	GG	AG	CT	GA	AG	AG	1500					
1501	GG	CT	CC	CC	CT	TG	CG	CT	TG	CG	CA	CC	CT	T	A	AG	GG	GG	GG	CG	CG	CA	AG	CG	GT	CG	1560				
1501	GG	CT	CC	CC	CT	TG	CG	CT	TG	CG	CA	CC	CT	T	A	AG	GG	GG	GG	CG	CG	CA	AG	CG	GT	CG	1560				
1561	AG	CC	CA	CC	CG	CA	CC	AA	AG	CT	TG	CT	GG	CA	AG	GT	T	CG	GG	CG	CT	CG	CC	CT	GT	CG	1620				
1561	AG	CC	CA	CC	CG	CA	CC	AA	AG	CT	TG	CT	GG	CA	AG	GT	T	CG	GG	CG	CT	CG	CC	CT	GT	CG	1620				
1621	TG	GA	C	AG	T	C	AG	GG	CCCC	CA	AA	C	AC	CG	GT	GT	CC	CG	T	AG	CC	CG	CC	CA	AA	AG	CT	GG	AT	C	1680
1621	TG	GA	C	AG	T	C	AG	GG	CCCC	CA	AA	C	AC	CG	GT	GT	CC	CG	T	AG	CC	CG	CC	CA	AA	AG	CT	GG	AT	C	1680
1681	CA	AA	C	AG	CT	TA	AT	TG	CG	CA	CC	CG	CT	TG	AG	CG	CT	TG	CT	GG	GG	CG	CT								


```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-12

Query Match      97.0%; Score 1782.8; DB 19; Length 3554;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1824; Conservative 0; Mismatches 2; Indels 30; Gaps 1;

QY      1  TGGAGCAGCCCTGGGCCCCCATCGGTATGAGGGCCGGGAAGTGGGGGGGTTGGGAGCC 60
DB      1513 TGGGAGCAGCCCTGGGCCCCCATCGGTATGAGGGCCGGGAAGTGGGGCGGTTGGGAGCC 1572

QY      61  TCCGTGGCCCTGCCCCCATCTCCAGTCCCAGTCCCTGGCTCGGACAGATAGGCGGAGGCT 120
DB      1573 TCCGTGGCCCTGCCCCCATCTCCAGTCCCAGTCCCTGGCTCGGACAGATAGGCGGAGGCT 1632

QY      121 GTGCTGTCTCTTTCA-----GAAGACTTTCGGGCGCA 150
DB      1633 GTGCTGTCTCTTTTCAGAGTAGCAGCGTGGGAAGAGGAGTTTCTTGGGAAGACTTCCGGCGCA 1692

QY      151 CATTGGCATCCAGCATCTCAAGTGGGTGTGAGCCACGAGCTCCAGCAGAGAGCAGCA 210
DB      1693 CATTGGCATCCAGCATCTCAAGTGGGTGTGAGCCACGAGCTCCAGCAGAGAGCAGCA 1752

QY      211 GAAGCAAGCCAGGGACCCAGAGGAGGAGGCGCGGAGCAGCAGCTGAGCAGGCAAG 270
DB      1753 GAAGCAAGCCAGGGACCCAGAGGAGGAGGCGCGGAGCAGCAGCTGAGCAGGCAAG 1812

QY      271 ATGCTGAAATGTGAGGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGGT 330
DB      1813 ATGCTGAAATGTGAGGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGGT 1872

QY      331 GCAAGTGTGCCAGGCTTACTTGGGAGCTGGGAGCAGTGGAGCATGAGGACATCGACAG 390
DB      1873 GCAAGTGTGCCAGGCTTACTTGGGAGCTGGGAGCAGTGGAGCATGAGGACATCGACAG 1932

QY      391 ATCCCTGTGAGGACCTCACTGAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTCTCC 450
DB      1933 ATCCCTGTGAGGACCTCACTGAGGCGCAGTGGGAGGACCTGACCCAGCAGTACTCTCC 1992

QY      451 TCGTTTCATGCGCATGCTTTTCATCTCAAATTCAGAAATTAATTTTCGCAATGCCAGGCTC 510
DB      1993 TCGTTTCATGCGCATGCTTTTCATCTCAAATTCAGAAATTAATTTTCGCAATGCCAGGCTC 2052

QY      511 TGCTGAATAGAAATCAGTCTGTGAACCTCAGAGGACATTTGACGGGCTCCGGAAATCT 570
DB      2053 TGCTGAATAGAAATCAGTCTGTGAACCTCAGAGGACATTTGACGGGCTCCGGAAATCT 2112

QY      571 GGATTATAAGCCCGGCGCAAGTCCCGGGCGGAGACATAGTGTGCATGGACCGTGTGG 630
DB      2113 GGATTATAAGCCCGGCGCAAGTCCCGGGCGGAGACATAGTGTGCATGGACCGTGTGG 2172

QY      631 AGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGACAAAGTGGGTGG 690
DB      2173 AGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGACAAAGTGGGTGG 2232

QY      691 TCCAGAAGTATACATCGAGAGCGCGCTGCTATCTGTGACACCAAGTTTCGACATCAGACAGT 750
DB      2233 TCCAGAAGTATACATCGAGAGCGCGCTGCTATCTGTGACACCAAGTTTCGACATCAGACAGT 2292

QY      751 GGTTCCTCTGTCCGAGTGAACCCCTGACCATCTGGTTCTACAAGGAGAGTTACTTGC 810
DB      2293 GGTTCCTCTGTCCGAGTGAACCCCTGACCATCTGGTTCTACAAGGAGAGTTACTTGC 2352

QY      811 GGTTCCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACCTGTGCAACA 870
DB      2353 GGTTCCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGCGCCATCCACCTGTGCAACA 2412

QY      871 ACCCGCTCCAGAAGTACTTGAAGAATGATGTGGGCGGAGCCCGCTGCTGCCGCGACACA 930
DB      871 ACCCGCTCCAGAAGTACTTGAAGAATGATGTGGGCGGAGCCCGCTGCTGCCGCGACACA 930
```

```
DB      2413 ACGCCCTCCAGAAAGTACCTGAAGAATGATGTGGGCGGAGCCCTGTGCTGCCGCGACACA 2472
QY      931 ACATGTGGACACAGCACCAAGTTCCAGGAGTACTCTGCAGCGCCAGGCGCCTGTGGCGCGTGT 990
DB      2473 ACATGTGGACACAGCACCAAGTTCCAGGAGTACTCTGCAGCGCCAGGCGCCTGTGGCGCGTGT 2532
QY      991 GGGGAGCGCTCATCTATACCCGTCCATGAAGAAGCCCATGCCCAACGCGCATGAAGGTGGCCC 1050
DB      2533 GGGGAGCGCTCATCTATACCCGTCCATGAAGAAGCCCATGCCCAACGCGCATGAAGGTGGCCC 2592
QY      1051 AGGACCACTGTGAGCGCTCGAAGAAACAGCTTTGAGCTCTACGGGGTGACTTCGTCTCTTG 1110
DB      2593 AGGACCACTGTGAGCGCTCGAAGAAACAGCTTTGAGCTCTACGGGGTGACTTCGTCTCTTG 2652
QY      1111 GGAGGAGACTTTCAGGCGCTCGGCTGATCGAGATCAATTTCCAGCGCCCAACATGCAATGCAATGCA 1170
DB      2653 GGAGGAGACTTTCAGGCGCTCGGCTGATCGAGATCAATTTCCAGCGCCCAACATGCAATGCAATGCA 2712
QY      1171 CGCCGCTCAACGCGCCAGCTGTGTGCAAGGTGAGGAGACACCATCAAGGTGGCCGCTGG 1230
DB      2713 CGCCGCTCAACGCGCCAGCTGTGTGCAAGGTGAGGAGACACCATCAAGGTGGCCGCTGG 2772
QY      1231 ACGCAGAGCTGTGACATCGGCACTTCGAGCTCTGTGAGGCGAGCCGGTGTGAGCCGC 1290
DB      2773 ACGCAGAGCTGTGACATCGGCACTTCGAGCTCTGTGAGGCGAGCCGGTGTGAGCCGC 2832
QY      1291 CCCCATTCAAGCGGGCTCCGACCTCTCGTGGCGGGGCTCAGTGTGAGGAGAGCCAGGAGGC 1350
DB      2833 CCCCATTCAAGCGGGCTCCGACCTCTCGTGGCGGGGCTCAGTGTGAGGAGAGCCAGGAGGC 2892
QY      1351 AGGTGCTGCCCTCTGTGCAACCTCAAGGCTTCGGCTTCGGTGTGAGCGGCGAGCGCTGA 1410
DB      2893 AGGTGCTGCCCTCTGTGCAACCTCAAGGCTTCGGCTTCGGTGTGAGCGGCGAGCGCTGA 2952
QY      1411 AGGCAAGCGGGCCCTCGGCCATGCCAGACCTGCGCCAGGAGCCCCATCACCAGCTCTCC 1470
DB      2953 AGGCAAGCGGGCCCTCGGCCATGCCAGACCTGCGCCAGGAGCCCCATCACCAGCTCTCC 3012
QY      1471 ACGGGGACTTGGGACTTGAAGGAAGAGAGGGGCTCCCCCTGGCCTTGTGGCACCCCTTAA 1530
DB      3013 ACGGGGACTTGGGACTTGAAGGAAGAGAGGGGCTCCCCCTGGCCTTGTGGCACCCCTTAA 3072
QY      1531 GGGGGGACGCGAGAGCGGTGAGCGGCGACAGCCACCCGACCAAAGTGTCTGGGAAGG 1590
DB      3073 GGGGGGACGCGAGAGCGGTGAGCGGCGACAGCCACCCGACCAAAGTGTCTGGGAAGG 3132
QY      1591 TGGAGCTCCCGGCTCGCCCTGTCGCCAGTGGAGAGTCAGGCGCCCAACACCGGTGTCC 1650
DB      3133 TGGAGCTCCCGGCTCGCCCTGTCGCCAGTGGAGAGTCAGGCGCCCAACACCGGTGTCC 3192
QY      1651 CGGTAGCCAGCCCGCCAAAGCTGGGATCCAAACCCAGCTAAATGCGCACCCGCTGGAGC 1710
DB      3193 CGGTAGCCAGCCCGCCAAAGCTGGGATCCAAACCCAGCTAAATGAGCACCCGCTGGAGC 3252
QY      1711 CTGTGCTGGGGGCTTGAAGACAGCAGAGGGCGCGTGTGCTCCGCCCGCCGAGGAGAAAG 1770
DB      3253 CTGTGCTGGGGGCTTGAAGACAGCAGAGGGCGCGTGTGCTCCGCCCGCCGAGGAGAAAG 3312
QY      1771 GTTCATGACAGCGTCAGATTCTCTGCAGCAGGAGTACAGGTTGCGAGCCACTCTCCC 1826
DB      3313 GTTCATGACAGCGTCAGATTCTCTGCAGCAGGAGTACAGGTTGCGAGCCACTCTCCC 3368

RESULT 4
US-10-635-977-12
; Sequence 12, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
```



```
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-6335-977-12

Query Match      97.0%; Score 1782.8; DB 19; Length 3554;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1824; Conservative 0; Mismatches 2; Indels 30; Gaps 1;

Qy 1 TGGGAGCAGGCTCGGCGCCCATCGGCTATGAGGCGGGAGTGGGGCGGGTTGGGGAGCC 60
Db 1513 TGGGAGCAGGCTCGGCGCCCATCGGCTATGAGGCGGGAGTGGGGCGGGTTGGGGAGCC 1572

Qy 61 TCCGTGCGCCCTGGCGCCCATCTCCAGTCCCAAGTCTCGCTCGGACAGATAGGGCGAGGCT 120
Db 1573 TCCGTGCGCCCTGGCGCCCATCTCCAGTCCCAAGTCTCGCTCGGACAGATAGGGCGAGGCT 1632

Qy 121 GTGCTGTCTTTCA-----GAAGACTTCCGGCGCA 150
Db 1633 GTGCTGTCTTTTACAGAGTAGCAGCGTGGGAGAGAGGAGTTCTTGGAAAGACTTCCGGCGCA 1692

Qy 151 CCATGGCATCCAGCATCTCTCAAGTGGGTGGTACGCCACAGAGCTGCAGCAGGAGGAGCA 210
Db 1693 CCATGGCATCCAGCATCTCTCAAGTGGGTGGTACGCCACAGAGCTGCAGCAGGAGGAGCA 1752

Qy 211 GAAGCAAGCCAGGAGCAGAGGAGGAGGCGCGGAGCAGCGACTTGGAGCAGCAGGCAAG 270
Db 1753 GAAGCAAGCCAGGAGCAGAGGAGGAGGCGCGGAGCAGCGACTTGGAGCAGCAGGCAAG 1812

Qy 271 ATGCTGAATATGCTGAGGCAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGGACATCGCGT 330
Db 1813 ATGCTGAATATGCTGAGGCAAGCTCAGGGGCTCCCGGGGCGAGCTTGTGGACATCGCGT 1872

Qy 331 GCAAGGTGTGCAGGCTTACTCTGGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAG 390
Db 1873 GCAAGGTGTGCAGGCTTACTCTGGGCGAGCTGGAGCATGAGGACATCGACACGTCAGCAG 1932

Qy 391 ATGCCGTGGAGGACCTCACTAGGCCGAGTGGAGGACCTGACCCAGCAGTACTCTCCC 450
Db 1933 ATGCCGTGGAGGACCTCACTAGGCCGAGTGGAGGACCTGACCCAGCAGTACTCTCCC 1992

Qy 451 TCGTTTCATGGGATGCTTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGTGCAGGCTC 510
Db 1993 TCGTTTCATGGGATGCTTTTCATCTCCAAATTCAGAAATTTACTTTTCGAGTGCAGGCTC 2052

Qy 511 TGCTGAATAGAAATCAAGCTCTGTGAACCTCAGACGGACATTTGACGGGCTCCGGAAATCT 570
Db 2053 TGCTGAATAGAAATCAAGCTCTGTGAACCTCAGACGGACATTTGACGGGCTCCGGAAATCT 2112

Qy 571 GGATTATAAGCCCGCGCCCAAGTCCCGGGGCGAGACATAGTGTGATGAGACCGGTGG 630
Db 2113 GGATTATAAGCCCGCGCCCAAGTCCCGGGGCGAGACATAGTGTGATGAGACCGGTGG 2172

Qy 631 AGGAGATCTGGAGCTGGGAGCTGCAGACCAACCTCTTTCCAGGGAACAAGTGGGTGG 690
Db 2173 AGGAGATCTGGAGCTGGGAGCTGCAGACCAACCTCTTTCCAGGGAACAAGTGGGTGG 2232

Qy 691 TCCAGAAATATCATCGAGACCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT 750
Db 2233 TCCAGAAATATCATCGAGACCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT 2292

Qy 751 GGTTCCTCGTCAGGACTGGAAACCCCTGACCATCTGTTCTACAGGAGAGTACTTGC 810
Db 2293 GGTTCCTCGTCAGGACTGGAAACCCCTGACCATCTGTTCTACAGGAGAGTACTTGC 2352
```

RESULT 5

US-10-615-659-11

; Sequence 11, Application US/10615659

; Publication No. US20040157234A1

```
Qy 811 GGTTCCTCAACTCAGCGCTTCTCCTCGACAAGCTGGACAGCGCATCCACTGTGCAACA 870
Db 2353 GGTTCCTCAACTCAGCGCTTCTCCTCGACAAGCTGGACAGCGCATCCACTGTGCAACA 2412

Qy 871 ACGCCGTCCAGAAATGACTCTGAAGAAATGATGTGGCGCGCAGCCCCCTCTGCTGCCGACACA 930
Db 2413 ACGCCGTCCAGAAATGACTCTGAAGAAATGATGTGGCGCGCAGCCCCCTCTGCTGCCGACACA 2472

Qy 931 ACATGTGGACAGGACACAGGTTCCAGAGTACCTGTGACGCGCAGGGCGGTGGCGCGCTGT 990
Db 2473 ACATGTGGACAGGACACAGGTTCCAGAGTACCTGTGACGCGCAGGGCGGTGGCGCGCTGT 2532

Qy 991 GGGCGACGCTCATCTACCCGCTCCATGAAGAGGCCATCGCCCCAGCCCATGAAGGTGGCCCC 1050
Db 2533 GGGCGACGCTCATCTACCCGCTCCATGAAGAGGCCATCGCCCCAGCCCATGAAGGTGGCCCC 2592

Qy 1051 AGGACACGCTGGAGCTTCGCAAGAACAGCTTTTTCAGAGCTCTACGGGGCTGACTTTCGTCTCTTG 1110
Db 2593 AGGACACGCTGGAGCTTCGCAAGAACAGCTTTTTCAGAGCTCTACGGGGCTGACTTTCGTCTCTTG 2652

Qy 1111 GAGGGGACTTTCAGGGCTCTGAGCTGATCGAGATCAATTCAGCCCCACCATGCAACCCGTCCA 1170
Db 2653 GAGGGGACTTTCAGGGCTCTGAGCTGATCGAGATCAATTCAGCCCCACCATGCAACCCGTCCA 2712

Qy 1171 CGCGGCTCAGCGGCCAGCTGTGTGCAAGGTGCGAGGAGCACCATCAAGGTGGCGCGTGG 1230
Db 2713 CGCGGCTCAGCGGCCAGCTGTGTGCAAGGTGCGAGGAGCACCATCAAGGTGGCGCGTGG 2772

Qy 1231 ACCGCACTGTGATCATCGGCAACTTCAGAGCTCTGTGGAGGAGCAGCGGTGGTGGAGCGCG 1290
Db 2773 ACCGCACTGTGATCATCGGCAACTTCAGAGCTCTGTGGAGGAGCAGCGGTGGTGGAGCGCG 2832

Qy 1291 CCCCATTCAGCGGTCGAGCTCTGTGTCGCGGGCTGCTGAGGAGGAGGAGGAGGAGG 1350
Db 2833 CCCCATTCAGCGGTCGAGCTCTGTGTCGCGGGCTGCTGAGGAGGAGGAGGAGGAGG 2892

Qy 1351 AGGTGTGTCGCGCTGCTGCAACCTCAAGGCTTCGCGCTCGCTGTTGGAGCGCGCGCTGGA 1410
Db 2893 AGGTGTGTCGCGCTGCTGCAACCTCAAGGCTTCGCGCTCGCTGTTGGAGCGCGCGCTGGA 2952

Qy 1411 AGGCAAGGGGCGCTTCGCGCATGCCAGACCTCTCCAGGAGACCCCATATCACAGCTCTCC 1470
Db 2953 AGGCAAGGGGCGCTTCGCGCATGCCAGACCTCTCCAGGAGACCCCATATCACAGCTCTCC 3012

Qy 1471 AGCGGACTTGGGACTGGAAGAGAGAGGGGCTCCCTTCGCGCTTGTCTGGCACCCCTTAA 1530
Db 3013 AGCGGACTTGGGACTGGAAGAGAGAGGGGCTCCCTTCGCGCTTGTCTGGCACCCCTTAA 3072

Qy 1531 GGGGGGACGCGAGAGGCGTGGAGCGCGCACAGCCCCACCGCACCAAGCTGCTGGGAGG 1590
Db 3073 GGGGGGACGCGAGAGGCGTGGAGCGCGCACAGCCCCACCGCACCAAGCTGCTGGGAGG 3132

Qy 1591 TGGAGCTCCCGGCTGCGCTGTCGACAGTGTGAGCAGTCAAGGCGCCCAACACACCGGTGTC 1650
Db 3133 TGGAGCTCCCGGCTGCGCTGTCGACAGTGTGAGCAGTCAAGGCGCCCAACACACCGGTGTC 3192

Qy 1651 CCGTAGCCACGCGCCCAAAAGCTGGGATCCAAACAGACTAAATGCGCAACCCGCTGGAGC 1710
Db 3193 CCGTAGCCACGCGCCCAAAAGCTGGGATCCAAACAGACTAAATGCGCAACCCGCTGGAGC 3252

Qy 1711 CTGTGCTGGGGGCTTGAAGACAGCAGGCGCGCTGCTGCGCGCGCGCGGAGGAGAAAG 1770
Db 3253 CTGTGCTGGGGGCTTGAAGACAGCAGGCGCGCTGCTGCGCGCGCGCGGAGGAGAAAG 3312

Qy 1771 GTTCATGACAGGCTCAGATTTCTGTGACAGGAGTACAGGTTCGAGCACTCTCCC 1826
Db 3313 GTTCATGACAGGCTCAGATTTCTGTGACAGGAGTACAGGTTCGAGCACTCTCCC 3368
```


; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-11

Query Match 91.3%; Score 1678.8; DB 19; Length 3465;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1757; Conservative 0; Mismatches 2; Indels 67; Gaps 1;

```
QY      1  TGGGAGCAGCCCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGGCGGCTTGGGGAGCC 60
DB      1526 TGGGAGCAGCCCTGGGCCCCCATCGGCTATGAGGGCGGGAAGTGGGGCGGCTTGGGGAGCC 1585

QY      61  TCCGTGGCCCTGGGCCCATCTCCAGTCCCCCAGTCCCTGGGCTCGGACAGATAGGGCGGAGGCT 120
DB      1586 TCCGTGGCCCTGGGCCCATCTCCAGTCCCCCAGTCCCTGGGCTCGGACAGATAGGGCGGAGGCT 1645

QY      121 GTGCTGTCTTTTTCAGAAAGACTTCCGGCGCACCATATGGCATCCAGCATCTCAAGTGGGTGG 180
DB      1646 GTGCTGTCTTTTTCAGAAAGACTTCCGGCGCACCATATGGCATCCAGCATCTCAAGTGGGTGG 1705

QY      181 TCAGCCACACAGACTCGACAGAGCAGCAGAGAGCAAGCCAGGACCCAGGACCCAGAGGGAGG 240
DB      1706 TCAGCCACACAGACTCGACAGAGCAGCAGAGAGCAAGCCAGGACCCAGGACCCAGAGGGAGG 1765

QY      241 CCGGGAGCAGCGACCTTGAGCAGCAGGCAAGATCTGAAATATCTGAGGCAAAAGCTCAGGG 300
DB      1766 CCGGGAGCAGCGACCTTGAGCAGCAGGCAAGATCTGAAATATCTGAGGCAAAAGCTCAGGG 1825

QY      301 GCCTCCCGGGCAGCTTGTGGAATCCGCTGCAAGGTGTGCCAGGCTACCTAGGGGCAGC 360
DB      1826 GCCTCCCGGGCAGCTTGTGGAATCCGCTGCAAGGTGTGCCAGGCTACCTAGGGGCAGC 1885

QY      361 TGGAGCATGAGGACATCGACACCTCAGCAGATCCGCTGGAGGACCTCAGTGGCCGAGT 420
DB      1886 TGGAGCATGAGGACATCGACACCTCAGCAGATCCGCTGGAGGACCTCAGTGGCCGAGT 1945

QY      421 GGGAGGACCTTGACCCAGCAGTACTACTCCCTCGCTTCATGGCGATGCTTTTCATCTCCAAAT 480
DB      1946 GGGAGGACCTTGACCCAGCAGTACTACTCCCTCGCTTCATGGCGATGCTTTTCATCTCCAAAT 2005

QY      481 CAAGAAATTAATTTTCGCGATGCGAGGCTCTGCTGAATAGAATCAGCTCTGTGAACCCCTC 540
DB      2006 CAAGAAATTAATTTTCGCGATGCGAGGCTCTGCTGAATAGAATCAGCTCTGTGAACCCCTC 2065

QY      541 AGACGGACATTCAGCGGCTCCGGAACATCTGGATTATAAGCCCGCGCCAGTCCCGGG 600
DB      2066 AGACGGACATTCAGCGGCTCCGGAACATCTGGATTATAAGCCCGCGCCAGTCCCGGG 2125

QY      601 GCCGAGACATAGTGTGCATGACCGGTGTGGAGGAGATCCTGGAGCTGGCGAGCTGCAGACC 660
DB      2126 GCCG----- 2129

QY      661 ACCCTCTTTCCAGGGACAACAATGGGTGTTCAGAAAGTACATCGAGACCGCGCTGCTCA 720
DB      2130 -----AGGGACAACAAGTGGGTGTTCAGAAAGTACATCGAGACCGCGCTGCTCA 2178

QY      721 TCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCTCGTACGGACTGGACCCCTCGA 780
DB      2179 TCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCTCGTACGGACTGGACCCCTCGA 2238
```

```
QY      781 CCATCTGGTTCTACAAGGAGAGTACTTGTGCGTTCTCAACTCAGCGCTTCTCCCTGGACA 840
DB      2239 CCATCTGGTTCTACAAGGAGAGTACTTGTGCGTTCTCAACTCAGCGCTTCTCCCTGGACA 2298

QY      841 AGCTGGACAGCGGCATCCACCTTGTGCAACAACCGCGTCCAGAACTACCTGAAGAATGATG 900
DB      2299 AGCTGGACAGCGGCATCCACCTTGTGCAACAACCGCGTCCAGAACTACCTGAAGAATGATG 2358

QY      901 TGGGGCGCAGCGCCCTGCTGCCCGGCACACAATGTGGACACGACCCAGGCTTCCAGGAGT 960
DB      2359 TGGGGCGCAGCGCCCTGCTGCCCGGCACACAATGTGGACACGACCCAGGCTTCCAGGAGT 2418

QY      961 ACCTGCAGCGCCAGGGCCCGTGGCGCGTGTGGGGCAGCGTCACTACCCGTCATGAAGA 1020
DB      2419 ACCTGCAGCGCCAGGGCCCGTGGCGCGTGTGGGGCAGCGTCACTACCCGTCATGAAGA 2478

QY      1021 AGGCCATCGCCACAGGCATGAAGGTGGCCACGACCAACGTGGAGCTCCCAAGAACAGCT 1080
DB      2479 AGGCCATCGCCACAGGCATGAAGGTGGCCACGACCAACGTGGAGCTCCCAAGAACAGCT 2538

QY      1081 TTGAGCTCTACGGGGCTGACTTTCGTCCTTGGAGGAGACTTCAGGGCCCTGGCTGATCGAGA 1140
DB      2539 TTGAGCTCTACGGGGCTGACTTTCGTCCTTGGAGGAGACTTCAGGGCCCTGGCTGATCGAGA 2598

QY      1141 TCAATTCACGCCCCACCATGCAACCGGTCCACGCCCGTCAAGCCCGCAGCTGTGTGCAAG 1200
DB      2599 TCAATTCACGCCCCACCATGCAACCGGTCCACGCCCGTCAAGCCCGCAGCTGTGTGCAAG 2658

QY      1201 TGCAGGAGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGC 1260
DB      2659 TGCAGGAGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGC 2718

QY      1261 TCTGTGGAGGACGCGGTGGTTGAGCGCCGCCCATTCAGCGGGTCCGACCTCTGCGTGG 1320
DB      2719 TCTGTGGAGGACGCGGTGGTTGAGCGCCGCCCATTCAGCGGGTCCGACCTCTGCGTGG 2778

QY      1321 CGGGCGTCAGTGTGAGGAGAGCCAGGAGCAGTGTCTGCCCTGTGCAACCTCAAGGCT 1380
DB      2779 CGGGCGTCAGTGTGAGGAGAGCCAGGAGCAGTGTCTGCCCTGTGCAACCTCAAGGCT 2838

QY      1381 CGGGCTCGCTGTGTGACGCGCAGCGCTGAAGGACGCGGGCCCCCTCGGCCATGCCAGACC 1440
DB      2839 CGGGCTCGCTGTGTGACGCGCAGCGCTGAAGGACGCGGGCCCCCTCGGCCATGCCAGACC 2898

QY      1441 CTGCCCAGGGACCCCATCACCAGCTCTCCAGCGGACTTGGGACTGAAGGAAGAGAAGG 1500
DB      2899 CTGCCCAGGGACCCCATCACCAGCTCTCCAGCGGACTTGGGACTGAAGGAAGAGAAGG 2958

QY      1501 GGCTCCCCCTGGCTTGTGTGGCAACCCTTAAGGGGGCAGCCGAGAGCGGTGGAGCGGCAC 1560
DB      2959 GGCTCCCCCTGGCTTGTGTGGCAACCCTTAAGGGGGCAGCCGAGAGCGGTGGAGCGGCAC 3018

QY      1561 AGCCACACCGGACCAAGCTGTGGGAGGTGGAGCTCCGGGCTGCCCTGTGCCACG 1620
DB      3019 AGCCACACCGGACCAAGCTGTGGGAGGTGGAGCTCCGGGCTGCCCTGTGCCACG 3078

QY      1621 TGGACAGTCAGGCCCCCAACACCGGTGTCCCGTAGCCAGCCCGCAAAAGCTGGGATC 1680
DB      3079 TGGACAGTCAGGCCCCCAACACCGGTGTCCCGTAGCCAGCCCGCAAAAGCTGGGATC 3138

QY      1681 CAAACCAGCTAAATGCGCACCCGCTGGAGCTGTGTCTGCGGGGCTGAAGA CAGCAGAGG 1740
DB      3139 CAAACCAGCTAAATGAGCACCCGCTGGAGCTGTGTCTGCGGAGCTGTGAAGACAGCAGAGG 3198

QY      1741 GCGCGCTGGCTCCGCGCCCGGAGAAAGGTTTCATGACAGCGTCAGATTCTCTCGAGCA 1800
DB      3199 GCGCGCTGGCTCCGCGCCCGGAGAAAGGTTTCATGACAGCGTCAGATTCTCTCGAGCA 3258

QY      1801 GGAGTACAGGTTGCAAGCCACTCTCCC 1826
DB      3259 GGAGTACAGGTTGCAAGCCACTCTCCC 3284
```


QY 1801 GGAGTACAGTTGACGCCACTCTCCC 1826
|
Db 3259 GGAGTACAGTTGACGCCACTCTCCC 3284
|
RESULT 7
US-10-615-659-9
; Sequence 9, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-9
Query Match 91.3%; Score 1677.2; DB 19; Length 1939;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1756; Conservative 0; Mismatches 3; Indels 67; Gaps 1;
QY 1 TGGGAGCAGCCCTGGGCCCCCATCGGCTATGAGGGCGGGAAGTGGGCGGTTTGGGGAGCC 60
Db 1 TGGGAGCAGCCCTGGGCCCCCATCGGCTATGAGGGCGGGAAGTGGGCGGTTTGGGGAGCC 60
QY 61 TCCGTGGCCCTGGCCCATCTCCAGTCCCGAGTCTTGGCTTCGACAGATAGGCGAGGCT 120
Db 61 TCCGTGGCCCTGGCCCATCTCCAGTCCCGAGTCTTGGCTTCGACAGATAGGCGAGGCT 120
QY 121 GTCTGTCTTTTTCAGAAAGACTTCCGGCGCACCATGGCATCCAGCATCTTCAAGTGGGTGG 180
Db 121 GTCTGTCTTTTTCAGAAAGACTTCCGGCGCACCATGGCATCCAGCATCTTCAAGTGGGTGG 180
QY 181 TCAGCCACACAGACTGCAGCAGGACGACGACGAGCAAGCCAGGACCCAGGAGGAGGAGG 240
Db 181 TCAGCCACACAGACTGCAGCAGGACGACGACGAGCAAGCCAGGACCCAGGAGGAGGAGG 240
QY 241 CCGGGAGCAGCAGCTTGCAGCAGGACGACGACGATGCTGAAAATGCTGAGGCAAAAGCTCAGGG 300
Db 241 CCGGGAGCAGCAGCTTGCAGCAGGACGACGACGATGCTGAAAATGCTGAGGCAAAAGCTCAGGG 300
QY 301 GCCTCCCGGGGAGCTTGTGGACATCGCGTGCAAGGTGTGCCAGGCTTACCTTGGGGCAGC 360
Db 301 GCCTCCCGGGGAGCTTGTGGACATCGCGTGCAAGGTGTGCCAGGCTTACCTTGGGGCAGC 360
QY 361 TGGAGCATGAGGACATCGACACGTACGACGATGCCGTGGAGGACCTTCACTGAGGCGGAGT 420
Db 361 TGGAGCATGAGGACATCGACACGTACGACGATGCCGTGGAGGACCTTCACTGAGGCGGAGT 420
QY 421 GGGAGGACCTTGACCCAGCAGTACTCTCCCTGTTTCAATGGCGATGCTTTTCACTCCAAATT 480
Db 421 GGGAGGACCTTGACCCAGCAGTACTCTCCCTGTTTCAATGGCGATGCTTTTCACTCCAAATT 480
QY 481 CAAGAAATTTACTTTTTCGAGTGCAGGCTCTGCTGAATAGAATTCAGTCTGTGAAACCTTC 540
Db 481 CAAGAAATTTACTTTTTCGAGTGCAGGCTCTGCTGAATAGAATTCAGTCTGTGAAACCTTC 540
QY 541 AGACGACATTGACGGGCTCCGGAAACATCTGGATTATAAGCCCGGCGCAAGTCCCGGG 600
Db 541 AGACGACATTGACGGGCTCCGGAAACATCTGGATTATAAGCCCGGCGCAAGTCCCGGG 600
QY 601 GCCGAGACATAGTGTGATGGACCCGTGTGGAGAGATCTTGGAGCTTGGCAGTGCAGACC 660
|

Db 601 GCGG----- 604
QY 661 ACCCTCTTTTCCAGGGACAAAGTGGGTGGTCCAGAAGTATCATCGAGACGCGCTCTCA 720
|
Db 605 -----AGGGACAAAGTGGGTGGTCCAGAAGTATCATCGAGACGCGCTCTCA 653
|
QY 721 TCTGTGACACCAAGTTTGCATCAGACAGTGGTTCCTCTCATCGGACTGGAACCCCTGA 780
|
Db 654 TCTGTGACACCAAGTTTGCATCAGACAGTGGTTCCTCTCATCGGACTGGAACCCCTGA 713
|
QY 781 CCATCTGGTTTCTACAGGAGAGTACTTCTCGGTTCCTCAACTCAGCGCTTCTCCCTCGACA 840
|
Db 714 CCATCTGGTTTCTACAGGAGAGTACTTCTCGGTTCCTCAACTCAGCGCTTCTCCCTCGACA 773
|
QY 841 AGCTGGACACGCGCATTCACCTGTGCAACAAGCCGCTCCAGAAGTACCTGAAGAAATGATG 900
|
Db 774 AGCTGGACACGCGCATTCACCTGTGCAACAACACCGTCCAGAAGTACCTGAAGAAATGATG 833
|
QY 901 TGGGCGGACAGCCCTCTGTGCGCGCACACAATGTGACACGACACAGGTTCCAGGAGT 960
|
Db 834 TGGGCGGACAGCCCTCTGTGCGCGCACACAATGTGACACGACACAGGTTCCAGGAGT 893
|
QY 961 ACCTGCAGCGCCAGGGCCGTGGCGCGTGTGGGGCAGCGTCACTTACCCGTCATCAAGA 1020
|
Db 894 ACCTGCAGCGCCAGGGCCGTGGCGCGTGTGGGGCAGCGTCACTTACCCGTCATCAAGA 953
|
QY 1021 AGGCCATCGCCACCGCATGAAGGTGGCCACAGAACACCGTGGAGCCTTCGAAAGACAGCT 1080
|
Db 954 AGGCCATCGCCACCGCATGAAGGTGGCCACAGAACACCGTGGAGCCTTCGAAAGACAGCT 1013
|
QY 1081 TTGAGCTCTACCGGGGTGACTTCTGTCCTTTGGGAGGAGCTTTCAGGCGCTGGCTGATCGAGA 1140
|
Db 1014 TTGAGCTCTACCGGGGTGACTTCTGTCCTTTGGGAGGAGCTTTCAGGCGCTGGCTGATCGAGA 1073
|
QY 1141 TCAATTTTCAGGCGCCACCATGACCGCTCCACCGCGTCAAGCGGCGAGCTGTGACACAG 1200
|
Db 1074 TCAATTTTCAGGCGCCACCATGACCGCTCCACCGCGTCAAGCGGCGAGCTGTGACACAG 1133
|
QY 1201 TGCAGGAGACACCATCAAGGTGGCGGTGGAACCGCAGCTGTGACATCGGCAACTTCGAGC 1260
|
Db 1134 TGCAGGAGACACCATCAAGGTGGCGGTGGAACCGCAGCTGTGACATCGGCAACTTCGAGC 1193
|
QY 1261 TCTGTGAGGAGCAGCGGTGGTGGAGCCGCCCTTTCAGCGGGTCCGACCTCTGGGTGG 1320
|
Db 1194 TCTGTGAGGAGCAGCGGTGGTGGAGCCGCCCTTTCAGCGGGTCCGACCTCTGGGTGG 1253
|
QY 1321 CGGGGTCTCATGTGTGAGGAGACCGAGGACAGTGTCTGCCCTTGCACACTCAAGGCT 1380
|
Db 1254 CGGGGTCTCATGTGTGAGGAGACCGAGGAGCAGGTGCTGCCCTTGCACACTCAAGGCT 1313
|
QY 1381 CGGCTCTGCTGTGGAACGCGCAGCGCTGAAGGCAAGGGCGGCGGCTCGGCGCATGCCAGACC 1440
|
Db 1314 CGGCTCTGCTGTGTGGAACGCGCAGCGCTGAAGGCAAGGGCGGCGGCTCGGCGCATGCCAGACC 1373
|
QY 1441 CTGCGCAGGAGACCCCATCAACAGCTCTCCAGCGGACCTTGGGACTGAAGGAGAGAAAG 1500
|
Db 1374 CTGCGCAGGAGACCCCATCAACAGCTCTCCAGCGGACCTTGGGACTGAAGGAGAGAAAG 1433
|
QY 1501 GGCTCCCGCTGGCTTGTGGCAACCTTAAAGGGGGGAGCGAGAGCGGTGGAGCCGAC 1560
|
Db 1434 GGCTCCCGCTGGCTTGTGGCAACCTTAAAGGGGGGAGCGAGAGCGGTGGAGCCGAC 1493
|
QY 1561 AGCCCAACCCGACCAAGCTGTGGAAGGTGGAGCTCCCGGCTGCGGCTGTGCGCAGC 1620
|
Db 1494 AGCCCAACCCGACCAAGCTGTGGAAGGTGGAGCTCCCGGCTGCGGCTGTGCGCAGC 1553
|
QY 1621 TGGACAGTTCAGGCGCCCAACACCGGTGTCCCGTGGCCAGCGCCCGCAAAAGTGGGATC 1680
|
Db 1554 TGGACAGTTCAGGCGCCCAACACCGGTGTCCCGTGGCCAGCGCCCGCAAAAGTGGGATC 1613
|
QY 1681 CAAACAGCTAAATGCGCACCCGCTGGAGCCTGTCTGGGGGCGCTTGAAGACAGCAGAGG 1740
|
Db 1614 CAAACAGCTAAATGAGCACCCCGCTGGAGCCTGTCTGGGAGCCTTGAAGACAGCAGAGG 1673
|

Qy 1741 GCGCGCTGCGTCCGCCCGCGGAGAAAGGTTTCATGACAGGTCAGATTCTCTGCAGCA 1800
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1674 GCGCGCTGCGTCCGCCCGCGGAGAAAGGTTTCATGACAGGTCAGATTCTCTGCAGCA 1733
Qy ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1801 GGAGTACAGGTTGACGACCTCTCC 1826
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1734 GGAGTACAGGTTGACGACCTCTCC 1759

RESULT 8
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match 91.3%; Score 1677.2; DB 19; Length 1939;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1756; Conservative 0; Mismatches 3; Indels 67; Gaps 1;

Qy 1 TGGGAGCAGGCTGGGCCCCCATCGGCTATGAGGGCGGGAAGTGGGGCGGGTGGGGAGCC 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 TGGGAGCAGGCTGGGCCCCCATCGGCTATGAGGGCGGGAAGTGGGGCGGGTGGGGAGCC 60

Qy 61 TCCGTGCGCTGGGCCCATCTCCAGTCCCGATCGGCTCGGACAGATAGGGCGAGGCT 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 TCCGTGCGCTGGGCCCATCTCCAGTCCCGATCGGCTCGGACAGATAGGGCGAGGCT 120

Qy 121 GTGCTGCTCTTTCAGAAAGCTTCGGGGCCACCATGGCATCCAGCATCTCAAGTGGGTGG 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 GTGCTGCTCTTTCAGAAAGCTTCGGGGCCACCATGGCATCCAGCATCTCAAGTGGGTGG 180

Qy 181 TCAGCCACCAAGCTGACGACGAGGACGACAGCAAGCCAGGGACAGAGGAGGAGG 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 TCAGCCACCAAGCTGACGACGAGGACGACAGCAAGCCAGGGACAGAGGAGGAGG 240

Qy 241 CCGGGACGAGGCTTCAGCAGCAGGACGAGATGCTGAAATGCTGAGGCAAAAGCTCAGGG 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 CCGGGACGAGGCTTCAGCAGCAGGACGAGATGCTGAAATGCTGAGGCAAAAGCTCAGGG 300

Qy 301 GCCTCCCGGGGACGCTTGTGACATCGCGTGCAGAGTGTGCCAGGCTTACCTGGGGCAGC 360
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 GCCTCCCGGGGACGCTTGTGACATCGCGTGCAGAGTGTGCCAGGCTTACCTGGGGCAGC 360

Qy 361 TGGAGCATGAGGACATCGACATGACAGATGCGTGGAGGACCTCACTGAGGCGGAGT 420
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 TGGAGCATGAGGACATCGACATGACAGATGCGTGGAGGACCTCACTGAGGCGGAGT 420

Qy 421 GGGAGGACCTGACCCAGCAGTACTCTCCCTCGTTTCATGCGATGCTTTTCAAT 480
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
421 GGGAGGACCTGACCCAGCAGTACTCTCCCTCGTTTCATGCGATGCTTTTCAAT 480

Qy 481 CAAGAAATTACTTTTCGAGTGGCAGGCTCTGCTGAATAGAAATCACGCTCTGTAACCCCTC 540
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
481 CAAGAAATTACTTTTCGAGTGGCAGGCTCTGCTGAATAGAAATCACGCTCTGTAACCCCTC 540

Qy 541 AGACGGACATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
541 AGACGGACATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600

Qy 601 GCGGACAGATAGTGTGATGGAACCGTGTGAGGAGATCTTGGAGCTGGCAGTGCAGACC 660
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
601 GCGG-----

Qy 661 ACCTCTTTCCAGGGAACAAGTGGGTGTCAGAAAGTACATCGAGACGCGCGTGTCTCA 720
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
605 -----AGGGAACAAGTGGGTGTCAGAAAGTACATCGAGACGCGCGTGTCTCA 653

Qy 721 TCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTGA 780
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
654 TCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTGA 713

Qy 781 CCATCTGGTTCACAAAGGAGGATTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGACA 840
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
714 CCATCTGGTTCACAAAGGAGGATTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTGGACA 773

Qy 841 AGCTGACAGCGGCTACCTCTGTGCAACAACGCGCTTCAGAAAGTACCTGAAAGTATG 900
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
774 AGCTGACAGCGGCTACCTCTGTGCAACAACGCGCTTCAGAAAGTACCTGAAAGTATG 833

Qy 901 TGGGCGCAGCGCCCTGCTGCCCGCACACAATGTGGAACGACAGCTTCCAGGAGT 960
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
834 TGGGCGCAGCGCCCTGCTGCCCGCACACAATGTGGAACGACAGCTTCCAGGAGT 893

Qy 961 ACCTGACGCGCAGGCGCGTGTGGGCGAGCGTCATCTACCCGTCATTAAGA 1020
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
894 ACCTGACGCGCAGGCGCGTGTGGGCGAGCGTCATCTACCCGTCATTAAGA 953

Qy 1021 AGGCAATGCGCCACGCGCATGAAGTGGCCGACAGACCAAGTGGAGCTTCGCAAGAACAGCT 1080
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
954 AGGCAATGCGCCACGCGCATGAAGTGGCCGACAGACCAAGTGGAGCTTCGCAAGAACAGCT 1013

Qy 1081 TTGAGCTCTACGCGGCTGACTTCGTCCTTGGGAGGAGTTCAGGCGCTGCTGATCCAGA 1140
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1014 TTGAGCTCTACGCGGCTGACTTCGTCCTTGGGAGGAGTTCAGGCGCTGCTGATCCAGA 1073

Qy 1141 TCAATTCAGCCCCACCATGACACCGCTCCACGCGGTCACGCGCCAGCTGTGTGCACAGG 1200
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1074 TCAATTCAGCCCCACCATGACACCGCTCCACGCGGTCACGCGCCAGCTGTGTGCACAGG 1133

Qy 1201 TGACGAGGACACCATCAAGTGGCGGTGACACGCAAGCTGTGACATCGGCAACTTCAGC 1260
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1134 TGACGAGGACACCATCAAGTGGCGGTGACACGCAAGCTGTGACATCGGCAACTTCAGC 1193

Qy 1261 TCCTGTGGAGGACGCGGTGTTGAGCGCGCCCATTTACGCGGGTCCGAGCTCTGCGTGG 1320
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1194 TCCTGTGGAGGACGCGGTGTTGAGCGCGCCCATTTACGCGGGTCCGAGCTCTGCGTGG 1253

Qy 1321 CGGCGCTCAGTGTGAGGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1254 CGGCGCTCAGTGTGAGGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1313

Qy 1381 CGGCGCTCAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1314 CGGCGCTCAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1373

Qy 1441 CTGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1374 CTGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1433

Qy 1501 GGGTCCCGCTTGGCTGCTGGGACCTTAAAGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1434 GGGTCCCGCTTGGCTGCTGGGACCTTAAAGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1493

Qy 1561 AGCCCAACCGGACCAAGAGTGTGGGAAGTGTGGAGCTCCCGGCTGCGCCCTGTGCGCACG 1620
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1494 AGCCCAACCGGACCAAGAGTGTGGGAAGTGTGGAGCTCCCGGCTGCGCCCTGTGCGCACG 1553

Qy 1621 TGGACAGTACGGGCCCCAACAACCGGTGTCCCGTAGCCCGAGCCCGGCAAGAGTGGGATC 1680

Db	1554	TGGACAGCTAGGCCCCAAACACCGGTGTCCCGTAGCCGACCGCCGCAAAAGCTGGGATC	1613
Qy	1681	CAAAACCAAGCTAAATGCGCACCCGCTGAGAGCCTGTGCTGCGGGGCTTGAAGACACAGCAGG	1740
Db	1614	CAAAACCAAGCTAAATAGACACCCGCTGAGAGCCTGTGCTGCGAGCTGAAGACAGCAGG	1673
Qy	1741	GCGCGCTGCGTTCGCGCCGCCCGAGGAGAAAGTTTCATCACACGCTCAGATTCTCTGCAGCA	1800
Db	1674	GCGCGCTGCGTTCGCGCCGCCCGAGGAGAAAGTTTCATCACACGCTCAGATTCTCTGCAGCA	1733
Qy	1801	GGAGTACAGCTTGACAGCCACTCTCCC	1826
Db	1734	GGAGTACAGCTTGACAGCCACTCTCCC	1759

RESULT 9

US-10-615-659-10

; Sequence 10, Application US/10615659

; Publication No. US20040157234A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN

; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42

; FILE REFERENCE: D0283 NP

; CURRENT APPLICATION NUMBER: US/10/615,659

; CURRENT FILING DATE: 2003-07-09

; PRIOR APPLICATION NUMBER: U.S. 60/394,725

; PRIOR FILING DATE: 2002-07-09

; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 10

; LENGTH: 1859

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-615-659-10

QY	615	TG	CATGGA	CCGTGTGGAGGAGATCTCTGGAGCTGGCGAGCTGCGAGACCA	CCCTCTTTTCCAGG	674
Db	466	TG	CATGGA	CCGTGTGGAGGAGATCTCTGGAGCTGGCGAGCTGCGAGACCA	CCCTCTTTTCCAGG	525
QY	675	GACA	CAAGTGGGTGGTCCAGAAGTACATCGAGACCGCCGCTGCTCATCTGTGACACCAAG	734		
Db	526	GACA	CAAGTGGGTGGTCCAGAAGTACATCGAGACCGCCGCTGCTCATCTGTGACACCAAG	585		
QY	735	TT	CGACATCAGACAGTGGTTCTCTCGTACGGAGCTGGAA	CCCCCTGACCATCTCTGGTCTCTAC	794	
Db	586	TT	CGACATCAGACAGTGGTTCTCTCGTACGGAGCTGGAA	CCCCCTGACCATCTCTGGTCTCTAC	645	
QY	795	AAGGAGAGTTACTTTGGGGTTCTCAACTCAGCGCTTCTCCCTTGGACAAGCTGGACGCGCC	854			
Db	646	AAGGAGAGTTACTTTGGGGTTCTCAACTCAGCGCTTCTCCCTTGGACAAGCTGGACGCGCC	705			
QY	855	AT	CCACTGTGCAACAAACCGCGTCCAGAAGTACCTGAAGAA	TGATGTGGCCCGCAGCCCC	914	
Db	706	AT	CCACTGTGCAACAAACCGCGTCCAGAAGTACCTGAAGAA	TGATGTGGCCCGCAGCCCC	765	
QY	915	CT	GCTCCCGCGCACACAACATGTGGACACGACACAGAGTTC	CAGGAGTACCTGCGAGCGCCAG	974	
Db	766	CT	GCTCCCGCGCACACAACATGTGGACACGACACAGAGTTC	CAGGAGTACCTGCGAGCGCCAG	825	
QY	975	GG	CGTGGCGCCGTGTGGGCGACGCTCATCTACCCCGTCCATGAAGAGG	CCATCGCCCCAC	1034	
Db	826	GG	CGTGGCGCCGTGTGGGCGACGCTCATCTACCCCGTCCATGAAGAGG	CCATCGCCCCAC	885	
QY	1035	GC	ATGAAGGTGGCCACAGACACAGGTGGAGCCTTCGCAAGAA	CAGCTTTGAGCTCTACGGG	1094	
Db	886	GC	ATGAAGGTGGCCACAGACACAGGTGGAGCCTTCGCAAGAA	CAGCTTTGAGCTCTACGGG	945	
QY	1095	GCT	GACTTTCGTCTTGGAGGGAGCTTTCAGGCGCTGCTGATCC	GAGATCAATTCAGGCCCC	1154	
Db	946	GCT	GACTTTCGTCTTGGAGGGAGCTTTCAGGCGCTGCTGATCC	GAGATCAATTCAGGCCCC	1005	
QY	1155	AC	CATGACCCCGTCCACGCGCGGTTCAGCGGCCAGCTGTGTG	CACAGGTGCGAGGAGCACCC	1214	
Db	1006	AC	CATGACCCCGTCCACGCGCGGTTCAGCGGCCAGCTGTGTG	CACAGGTGCGAGGAGCACCC	1065	
QY	1215	AT	CAAGGTGGCCGTGGAACCGCAGCTGTGACATCGGCAAT	TCGGCACTTGTGGAGGCGAG	1274	
Db	1066	AT	CAAGGTGGCCGTGGAACCGCAGCTGTGACATCGGCAAT	TCGGCACTTGTGGAGGCGAG	1121	
QY	1275	CG	GTGTTGAGCGCGCCCGCATTTACGCGGCTCGACCTCTG	CGTGGCGGCGTCAAGTG	1334	
Db	1122	CG	GTGTTGAGCGCGCCCGCATTTACGCGGCTCGACCTCTG	CGTGGCGGCGTCAAGTG	1181	
QY	1335	AG	GAGACCGAGGAGGAGGTGTCTGCCCGTCTGCAACCTC	CAAGGCTCTCGGCTCGCTGTTG	1394	
Db	1182	AG	GAGACCGAGGAGGAGGTGTCTGCCCGTCTGCAACCTC	CAAGGCTCTCGGCTCGCTGTTG	1241	
QY	1395	GAC	CGCAGCCGCTGAAGGCA	CGGGCGCTCTGGCCATGCGCAGACCTTGGCCAGGACCC	1454	
Db	1242	GAC	CGCAGCCGCTGAAGGCA	CGGGCGCTCTGGCCATGCGCAGACCTTGGCCAGGACCC	1301	
QY	1455	CC	ATCAACCAAGCTCTCAGCGGAGCTTTGGAGACTGAAG	AGAGAGGGGCTCCCGCTGGCC	1514	
Db	1302	CC	ATCAACCAAGCTCTCAGCGGAGCTTTGGAGACTGAAG	AGAGAGGGGCTCCCGCTGGCC	1361	
QY	1515	TT	GCTGGCACCCCTTAAGGGGGG	CAGCCGAGCGGTGGAGCCGCGCACAGCCCGCCAGCC	1574	
Db	1362	TT	GCTGGCACCCCTTAAGGGGGG	CAGCCGAGCGGTGGAGCCGCGCACAGCCCGCCAGCC	1421	
QY	1575	AA	AGCTGTCTGGGAAGGTGAGCTCTCCGGCGTCCCTCTG	CGCACTGTGACAGTCAAGCC	1634	
Db	1422	AA	AGCTGTCTGGGAAGGTGAGCTCTCCGGCGTCCCTCTG	CGCACTGTGACAGTCAAGCC	1481	
QY	1635	CC	AAACACACCGGTGTCCCGTACGCCACCGCCCAAGAG	CTGGATCCAAACCAAGCTAAAT	1694	
Db	1482	CC	AAACACACCGGTGTCCCGTACGCCACCGCCCAAGAG	CTGGATCCAAACCAAGCTAAAT	1541	
QY	1695	GG	CGACCCCGCTGGAGCCTGTGCTCGGGGCGCTCAAG	ACACAGAGAGGGCGCGCTGCGCTCG	1754	

Db	1542	GAGCACCCGCTGGAGCCCTGTGCTCGGAGCCTGAAGA	CAGCAGAGGGCGCGCTGCCTGCGCCG	1601
Qy	1755	CCGCCCGGAGGAAAAGGTTTCATGACAGCGT	CAGATTCTCTGCAGCAGGAGTACAGGTTGC	1814
Db	1602	CCGCCCGGAGGAAAAGGTTTCATGACAGGTT	CAGATTCTCTGCAGCAGGAGTACAGGTTGC	1661
Qy	1815	AGCCACTCTCCC	1826	
Db	1662	AGCCACTCTCCC	1673	
RESULT 10				
US-10-635-977-10				
; Sequence 10, Application US/10635977				
; Publication No. US20040171131A1				
; GENERAL INFORMATION:				
; APPLICANT: Bristol-Myers Squibb Company				
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN				
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42				
; FILE REFERENCE: D0283A CIP				
; CURRENT APPLICATION NUMBER: US/10/635,977				
; CURRENT FILING DATE: 2003-08-07				
; PRIOR APPLICATION NUMBER: U.S. 60/394,725				
; PRIOR FILING DATE: 2002-07-09				
; PRIOR APPLICATION NUMBER: U.S.10/615,659				
; PRIOR FILING DATE: 2003-07-09				
; NUMBER OF SEQ ID NOS: 103				
; SOFTWARE: PatentIn version 3.2				
; SEQ ID NO 10				
; LENGTH: 1859				
; TYPE: DNA				
; ORGANISM: Homo sapiens				
US-10-635-977-10				
Query Match 85.7%; Score 1574.8; DB 19; Length 1859;				
Best Local Similarity 97.1%; Pred. No. 0;				
Matches 1643; Conservative 0; Mismatches 2; Indels 47; Gaps 2;				
Qy	135	GAAGACTTCCGGCGCACCATGGCATCCAGCATCTCTCAAGTGGGTGGTCAAGCACCAGAGC	194	
Db	29	GAAGACTTCCGGCGCACCATGGCATCCAGCATCTCTCAAGTGGGTGGTCAAGCACCAGAGC	88	
Qy	195	TGCAGCAGGAGCAGCAGAGCAAGCCAGGGACCCAGAGGAGGAGGCCGGGAGCAGCGAC	254	
Db	89	TGCAGCAGGAGCAGCAGAGCAAGCCAGGGACCCAGAGGAGGAGGCCGGGAGCAGCGAC	148	
Qy	255	CTGAGCAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAGGGGCGCTCCCGGGGCGAG	314	
Db	149	CTGAGCAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAGGGGCGCTCCCGGGGCGAG	208	
Qy	315	CTTGTGACATCGCGTCACAGGTGTGCCAGGCTTACCTGGGGCAGCTGAGGAGCATGAGGAC	374	
Db	209	CTTGTGACATCGCGTCACAGGTGTGCCAGGCTTACCTGGGGCAGCTGAGGAGCATGAGGAC	268	
Qy	375	ATCGACACGTCAGCAGATGCGTGGAGGACCTCACTGAGGCCAGCTGGGAGGACCTGAGCC	434	
Db	269	ATCGACACGTCAGCAGATGCGTGGAGGACCTCACTGAGGCCAGCTGGGAGGACCTGAGCC	328	
Qy	435	CAGCAGTACTCTCCCTCGTTTCATGCGGATGCTTTTCATCTCCAATTCAAGAAATTACTTT	494	
Db	329	CAGCAGTACTCTCCCTCGTTTCATCTCCAATTCAAGAAATTACTTT	351	
Qy	495	TCGCAGTGGCCAGGCTCTGCTGTAATAGAAATCAAGTCTGTGAACCTTCAGACGGACNTGAC	554	
Db	352	-----TGCCAGGCTCTGCTGTAATAGAAATCAAGTCTGTGAACCTTCAGACGGACNTGAC	405	
Qy	555	GGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCCAAAGTCCCGGGGCGCAGACATAGTG	614	
Db	406	GGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCCAAAGTCCCGGGGCGCAGACATAGTG	465	
Qy	615	TGCATGGAACCGTGTGGAGGAGATCCTGGAGCTGGCAGCTCGACACCAACCTCTTTCCAGG	674	

466	TGCATGACCGTGTGGAGGAGATCCTCGAGCTGGCAGCTCGAGACACACCTCTTCTTCAGG	525
675	GACAAACAAGTGGGTGGTCCAGAAATCATCGAGACGCCGTGCTCATCTGTGACACCAAG	734
526	GACAAACAAGTGGGTGGTCCAGAAATCATCGAGACGCCGTGCTCATCTGTGACACCAAG	585
735	TTTGACATCAGACAGTGGTTCCTCGTACGGACTGGAAACCCCTGACCATCTGTTCTTAC	794
586	TTTGACATCAGACAGTGGTTCCTCGTACGGACTGGAAACCCCTGACCATCTGTTCTTAC	645
795	AAGGAGAGTTACTTTGCGGTTCTAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCC	854
646	AAGGAGAGTTACTTTGCGGTTCTAACTCAGCGCTTCTCCCTGGACAAGCTGGACAGCGCC	705
855	ATCCACCTGTGCACAAACGCCGTGCCAGAAGTACCTGAAGAAATGATGTGGGCCGACGCCCC	914
706	ATCCACCTGTGCACAAACGCCGTGCCAGAAGTACCTGAAGAAATGATGTGGGCCGACGCCCC	765
915	CTGCTGCCCGGACACAAACATGTGGACAGACACAGGTTCCAGGAGTACCTGCAGCGGCAG	974
766	CTGCTGCCCGGACACAAACATGTGGACAGACACAGGTTCCAGGAGTACCTGCAGCGGCAG	825
975	GGCGTGGCGCGTGTGGGCGACGCTCATCTACCCGTCCATGAAGAAGGCCATCGCCAC	1034
826	GGCGTGGCGCGTGTGGGCGACGCTCATCTACCCGTCCATGAAGAAGGCCATCGCCAC	885
1035	GCCATGAAGTGGGCCACAGGACCACTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGG	1094
886	GCCATGAAGTGGGCCACAGGACCACTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGG	945
1095	GCTGACTTCGTTCTTTGGGAGGGGACTTCAGGCGCTGTGTGATCGAGATCAATTCCAGCCCC	1154
946	GCTGACTTCGTTCTTTGGGAGGGGACTTCAGGCGCTGTGTGATCGAGATCAATTCCAGCCCC	1005
1155	ACCATGACCCCGTCCAGCGCGGTACGGCCCGACTGTGTCGACAGGTGCGAGGAGACACC	1214
1006	ACCATGACCCCGTCCAGCGCGGTACGGCCCGACTGTGTCGACAGGTGCGAGGAGACACC	1065
1215	ATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGAG	1274
1066	ATCAAGGTGGCGGTGGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGAG	1125
1275	CCGGTGTGAGCGCGCCCAATTCAGCGGGTTCGACCTCTGCTGTGGCGGGCGTCAGTGTG	1334
1122	CCGGTGTGAGCGCGCCCAATTCAGCGGGTTCGACCTCTGCTGTGGCGGGCGTCAGTGTG	1181
1335	AGGAGACGACGAGGACAGGTGCTGCCCGTCTGCAACTCAAGCGCTCGGCGCTCGCTGTG	1394
1182	AGGAGACGACGAGGACAGGTGCTGCCCGTCTGCAACTCAAGCGCTCGGCGCTCGCTGTG	1241
1395	GACGCGACGCGCTGAAGGACACGGGGCCCTCGGCCATGCGACACCTGCGCAGGACCC	1454
1242	GACGCGACGCGCTGAAGGACACGGGGCCCTCGGCCATGCGACACCTGCGCAGGACCC	1301
1455	CCATCACCACTCTCCACGGGACTTTGGGACTGAAGGAAGAGAGGGGCTCCCCCTGGCC	1514
1302	CCATCACCACTCTCCACGGGACTTTGGGACTGAAGGAAGAGAGGGGCTCCCCCTGGCC	1361
1515	TTGCTGACACCTTTAAGGGGGGACGCCGAGAGCGGTGGAGCGGCAGACGCCGACCC	1574
1362	TTGCTGACACCTTTAAGGGGGGACGCCGAGAGCGGTGGAGCGGCAGACGCCGACCC	1421
1575	AAAGCTGCTGGGAAGGTGGAGCTCCCGGCTGCCCTGTGCCACGTGGAAGTGGAGCGCC	1634
1422	AAAGCTGCTGGGAAGGTGGAGCTCCCGGCTGCCCTGTGCCACGTGGAAGTGGAGCGCC	1481
1635	CCAAACACCGGTGTCCCGGTAGCCGACGCCCGCCAAAAGCTGGGATCCAAACACAGCTAAAT	1694
1482	CCAAACACCGGTGTCCCGGTAGCCGACGCCCGCCAAAAGCTGGGATCCAAACACAGCTAAAT	1541
1695	GCGCACCCCGCTGAGACCTGTGCTGCGGGGCTTGAGACAGACAGGCGCGCTGCGTCCG	1754
1542	GAGCACCCCGCTGAGACCTGTGCTGCGGAGCGCTGAAGACAGCAGAGGCGCGCTGCGTCCG	1601

Qy	1755	CCGCCCGAGGAAAAGTTTCATGACAGCTCAGATTCTCTGCAGCAGGAGTACAGGTTGC	1814
Db	1602	CCGCCCGAGGAAAAGTTTCATGACAGCTCAGATTCTCTGCAGCAGGAGTACAGGTTGC	1661
Qy	1815	AGCCACTCTCCC	1826
Db	1662	AGCCACTCTCCC	1673

```

RESULT 11
US-10-615-659-3
; Sequence 3, Application US/10615659
; Publication No. US2004015734A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BG542
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-615-659-3

```

Query Match	39.5%	Score 726;	DB 19;	Length 726;
Best Local Similarity	100.0%;	Pred. No. 1.5e-189;		
Matches 726;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps 0;

549	Qy	ATTGACGGGCTCGGAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGGCCGAGAC	608
1	Db	ATTGACGGGCTCGGAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGGCCGAGAC	60
609	Qy	ATAGTGTGCATGGACCCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT	668
61	Db	ATAGTGTGCATGGACCCGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCAACCCCTCTT	120
669	Qy	TCCAGGGAACAACAGTGGTGGTCCAGAATACATCGAGAGCGCGCTGCTCATCTGTGCAC	728
121	Db	TCCAGGGAACAACAGTGGTGGTCCAGAATACATCGAGAGCGCGCTGCTCATCTGTGCAC	180
729	Qy	ACCAAGTTTGACATCAGACAGTGGTCTCGTCAACGGACTGGAAACCCCTGACCATCTGG	788
181	Db	ACCAAGTTTGACATCAGACAGTGGTCTCGTCAACGGACTGGAAACCCCTGACCATCTGG	240
789	Qy	TTCTCAAGGAGAGTTACTTGGGTTTCTAACTCAGCGCTTCTCCCTGGACAAGCTGGAC	848
241	Db	TTCTCAAGGAGAGTTACTTGGGTTTCTAACTCAGCGCTTCTCCCTGGACAAGCTGGAC	300
849	Qy	AGCGCCATCACCTGTGCACAACCGCGTCCAGAAGTACCTGGAAGATGATGTGGGCGCG	908
301	Db	AGCGCCATCACCTGTGCACAACCGCGTCCAGAAGTACCTGGAAGATGATGTGGGCGCG	360
909	Qy	AGCCCCCTGTGCCCGCACAAATGTGGACCAAGCACCGAGTTCCAGGAGTACCTGCAG	968
361	Db	AGCCCCCTGTGCCCGCACAAATGTGGACCAAGCACCGAGTTCCAGGAGTACCTGCAG	420
969	Qy	CGCAGGGCCGTGGCGCGGTGTGGGCGAGCGTCACTTACCCGTCCATGAGAGAGGCCATC	1028
421	Db	CGCAGGGCCGTGGCGCGGTGTGGGCGAGCGTCACTTACCCGTCCATGAGAGAGGCCATC	480
1029	Qy	GCCCAAGCCATGAAGGTGGCCCAAGGACACGTGGAGCCTCGCAAGAAACAGCTTTGAGCTC	1088
481	Db	GCCCAAGCCATGAAGGTGGCCCAAGGACACGTGGAGCCTCGCAAGAAACAGCTTTGAGCTC	540

Qy	1089	TACGGGGCTGACATTCGTCCTTGGAGGGACTTCAGGCCCTCGGTGATCGAGATCAATTC	1144
Db	541	TACGGGGCTGACATTCGTCCTTGGAGGGACTTCAGGCCCTCGGTGATCGAGATCAATTC	600
Qy	1149	AGCCCCACCATGCACCCGCTCCAGCCCGTCA CGGCCCGAGCTGTGTGCACAGGTGCAGGAG	1208
Db	601	AGCCCCACCATGCACCCGCTCCAGCCCGTCA CGGCCCGAGCTGTGTGCACAGGTGCAGGAG	660
Qy	1209	GACACCATCAAGGTGGCCGTGGAGCCGACGCTGTGATCATCGGCAACTTCGAGCTCTGTGG	1268
Db	661	GACACCATCAAGGTGGCCGTGGAGCCGACGCTGTGATCATCGGCAACTTCGAGCTCTGTGG	720
Qy	1269	AGGCAG 1274	
Db	721	AGGCAG 726	
RESULT 12			
US-10-635-977-3			
; Sequence 3, Application US/10635977			
; Publication No. US20040171191A1			
; GENERAL INFORMATION:			
; APPLICANT: Bristol-Myers Squibb Company			
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN			
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42			
; FILE REFERENCE: D0283A CIP			
; CURRENT APPLICATION NUMBER: US/10/635,977			
; CURRENT FILING DATE: 2003-08-07			
; PRIOR APPLICATION NUMBER: U.S. 60/394,725			
; PRIOR FILING DATE: 2002-07-09			
; PRIOR APPLICATION NUMBER: U.S.10/615,659			
; PRIOR FILING DATE: 2003-07-09			
; NUMBER OF SEQ ID NOS: 103			
; SOFTWARE: Patent In version 3.2			
; SEQ ID NO 3			
; LENGTH: 726			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (1)...(726)			
US-10-635-977-3			
Query Match 39.5%; Score 726; DB 19; Length 726;			
Best Local Similarity 100.0%; Pred. No. 1.5e-189;			
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	549	ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCCAAGTCCCGGGGCCGAGAC	608
Db	1	ATTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCCAAGTCCCGGGGCCGAGAC	60
Qy	609	ATAGTGTGCATGACCCCTGTGTGGAGAGATCCTGGAGTGGCAGTGCAGACCAACCTCTTT	669
Db	61	ATAGTGTGCATGACCCCTGTGTGGAGAGATCCTGGAGTGGCAGTGCAGACCAACCTCTTT	120
Qy	669	TCCAGGACCAACAGTGGGTGGTCCAGAGTATATCGAGAGCGCGTGTCTATCTGTGAC	728
Db	121	TCCAGGACCAACAGTGGGTGGTCCAGAGTATATCGAGAGCGCGTGTCTATCTGTGAC	180
Qy	729	ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTGACCATCTGG	788
Db	181	ACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCA CGGACTGGAAACCCCTGACCATCTGG	240
Qy	789	TTCTCAAGGAGAGTTACTTTCGCGTTCTCAACTCAGCGCTTCTTCCTCTGGACAAGCTGGAC	848
Db	241	TTCTCAAGGAGAGTTACTTTCGCGTTCTCAACTCAGCGCTTCTTCCTCTGGACAAGCTGGAC	300
Qy	849	AGGCGCATCCACTGTGCAACACCGCGTCCAGAGTACTCTGAGAGATGATGTGGGCGGC	908
Db	301	AGGCGCATCCACTGTGCAACACCGCGTCCAGAGTACTCTGAGAGATGATGTGGGCGGC	360
Qy	909	AGGCCCCGTGTGCCCGCACCAACATGTGGACCAAGTTCACAGGTTTCCAGGAGTACCTCGAC	968

Db 361 AGCCCCCTGCTGCCCGCACACATGTGACACAGACACAGGTTCCAGAGTACTCTGCAG 420
Qy 969 CGCCAGGGCCGTGGCGCGGTGTGGGCGAGCGTCACTTACCCGTCCATGAAGAAGGCCATC 1028
Db 421 CGCCAGGGCCGTGGCGCGGTGTGGGCGAGCGTCACTTACCCGTCCATGAAGAAGGCCATC 480
Qy 1029 GCCCAGCCATGAAGTGGCCCGCAGGACCAAGTGGAGCGCTCGCAAGACAGCTTTGAGCTC 1088
Db 481 GCCCAGCCATGAAGTGGCCCGCAGGACCAAGTGGAGCGCTCGCAAGACAGCTTTGAGCTC 540
Qy 1089 TAGCGGGCTGACTTTGCTTCTTGGGAGGAGCTTTCAGGCGCTCGGTGATCGAGATCAATTCC 1148
Db 541 TAGCGGGCTGACTTTGCTTCTTGGGAGGAGCTTTCAGGCGCTCGGTGATCGAGATCAATTCC 600
Qy 1149 AGCCCAACCATGACCCCGTCAAGCCCGTCAAGCCCGTCAAGCCCGTCAAGTGGCAGGAG 1208
Db 601 AGCCCAACCATGACCCCGTCAAGCCCGTCAAGCCCGTCAAGCCCGTCAAGTGGCAGGAG 660
Qy 1209 GACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCTGTGG 1268
Db 661 GACACCATCAAGTGGCGCGTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCTGTGG 720
Qy 1269 AGGCAG 1274
Db 721 AGGCAG 726

RESULT 13
US-10-723-860-631/c
; Sequence 631, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma. Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 631
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-631

Query Match 27.3%; Score 501; DB 20; Length 101270;
Best Local Similarity 100.0%; Pred. No. 3.2e-127;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1272 CAGCCGTGTGTAGCCGCCGCCCAATTACGGGGTTCGACCTCTCGTGGCGGGCGTCACT 1331
Db 43637 CAGCCGTGTGTAGCCGCCGCCCAATTACGGGGTTCGACCTCTCGTGGCGGGCGTCACT 43578
Qy 1332 GTGAGGAGACGAGGAGGAGGTGCTGCCCGTCTGCAACCTCAAGGCTCGGGCTCGCTG 1391
Db 43577 GTGAGGAGACGAGGAGGAGGTGCTGCCCGTCTGCAACCTCAAGGCTCGGGCTCGCTG 43518
Qy 1392 TTGAGCGCGCAGCGCTGAAGGACGCGGGGCCCTTCGGCCATGCGACACCTGCCAGGGA 1451
Db 43517 TTGAGCGCGCAGCGCTGAAGGACGCGGGGCCCTTCGGCCATGCGACACCTGCCAGGGA 43458
Qy 1452 CCCCCATCACAGCTTCTCAGCGGGACTTGGGACTGAAGAAGAGAGGGGCTCCCGCTG 1511
Db 43457 CCCCCATCACAGCTTCTCAGCGGGACTTGGGACTGAAGAAGAGAGGGGCTCCCGCTG 43398
Qy 1512 GCCTTGTGACCCCTTAAGGGGGGCGAGCGAGCGGTGAGCCGCGCACAGCCACCCCG 1571

Db 43397 GCCTTGTGSCACCCCTTAAGGGGGGCGAGCGAGCGGTGGAGCCGCGCACAGCCACCCCG 43338
Qy 1572 ACCAAAGCTCTCGGAAGGTGGAGCTCCCGGCTTCCGCTTGTGCGCCAGCTGGACAGTCAG 1631
Db 43337 ACCAAAGCTCTCGGAAGGTGGAGCTCCCGGCTTCCGCTTGTGCGCCAGCTGGACAGTCAG 43278
Qy 1632 GCCCAAAACACACCGTGTCCCGGTAGCCAGCGCCGCAAAAGCTGGGATCCAAACCACTA 1691
Db 43277 GCCCAAAACACCGTGTCCCGGTAGCCAGCGCCGCAAAAGCTGGGATCCAAACCACTA 43218
Qy 1692 AATGCGACCCCGTGGAGCGCTGTGTCGGGGGCTTGAAGACAGCAGAGGCGCGCTCGT 1751
Db 43217 AATGCGACCCCGTGGAGCGCTGTGTCGGGGGCTTGAAGACAGCAGAGGCGCGCTCGT 43158
Qy 1752 CGCCCGCCCGGAGGAAAAGT 1772
Db 43157 CGCCCGCCCGGAGGAAAAGT 43137

RESULT 14
US-10-275-595A-39
; Sequence 39, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dying Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BATRA, Sajeev
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 2380
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 2156553CB1
US-10-275-595A-39

Query Match 23.5%; Score 432.8; DB 18; Length 2380;
Best Local Similarity 59.1%; Pred. No. 1e-108;
Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;

Qy 260 CAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCTCCCGGGGAGCTTG 318
Db 361 CAGGACACACAGCAGCGCCCAAGAAAACAGAGAAAACCCAGTGTGTGTTCCCGAGAGTTTG 420
Qy 319 TGGACATCCGTCAGAGTGTGCCAGGCTTACCTGGGGAGCTGGAGCATGAGGACATCG 378

Db 421 TGGATGAAGCTCTGTGTGGTGGAGGAGTACCTTAGCAACATTGGGCCCAATGGACATCG 480
Qy 379 ACACGTCAGCAGATGCCGTGGAGAGCCTCACTGAGGCCGAGTGGGAGGACCTGACCCAGC 438
Db 481 ACAAGGACCTGGAGGCCCGCTGTACTCTACCCCGAGGGCTGGTCCCTCTTCTCCAGC 540
Qy 439 AGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATATCTTTTCGC 498
Db 541 GCTACTACCAAGTGTCTCAGGAAGGGGCGAGAACTCAGGCACCTCGACACTCAGAGTCCAGC 600
Qy 499 AGTGCCAGGCTCTGCTGAATAGNATCACGTCCTGTGAACCTCAGACGGACATTTGACGGC 558
Db 601 GCTGTGAGGACATCTCTGACGAGCTCGAGCGCTGGTACCCCGAGTAGAGTAGGAAGGG 660
Qy 559 TCGGAACATCTGGATTATAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTTGCA 618
Db 661 ATCGCAACATCTGGATCGTAGCCAGGAGCCAGTCCCGTGGACGAGGCATCATGTGCA 720
Qy 619 TGAACCGTGTGGAGGAGATCCTGGAGCTGGAGCTGCAGACCAACCTCTTTTCAGGGACA 678
Db 721 TGAACCACTGGAGGAGATGCTGAGCTGGTGAACCGCAACCCCGTGGTGTGATGAAGGAG 780
Qy 679 CAAAGTGGTGTCCAGAGTATCATCGAGACCGCTGCTCATCTGTGACACCAAGTTCCG 738
Db 781 GCAAGTGGTGTGCGAAGTATATTGAGCGGCCCTCTCTCATCTTTTGGACCAAGTTTG 840
Qy 739 ACATCAGACAGTGGTTCTCTCGTCAAGGACCTGGAACCCCTGACCATCTGTTCTCAAGG 798
Db 841 ACCTCAGACAGTGGTTCTCTGTAACCTGACTGGAACCCACTTACCGTGTGTTCTACCGG 900
Qy 799 AGAGTTACTTGGCGTTCTCAACTCAGCGCTTCTCCCTGGAACGCTGGACAGCGCCATCC 858
Db 901 ACAGCTATATCCGCTTTTCCAGCGCAGCCCTTCTCCCTGAAGAACCTGGACAACCTCAGTGC 960
Qy 859 ACTGTGCAACACGCGTCCAGAGTACTGAGNATGATGTGGCCCGCAGCCCTCTGC 918
Db 961 ACTGTGCAACAACTCCATCCAGAGACCTGGAGAACTCATGCGCATCGGCATCTCACTGC 1020
Qy 919 TGCCCGCAACAACATGTGGACAGACCAAGTTCCAGAGTACCTGACGCGCCAGGGCC 978
Db 1021 TTCCGCGACACATGTGTTAGCCAGAGGTTCCAGGCCCACTGACGAGATGGGTG 1080
Qy 979 GTGGCCCGTGTGGGACAGGTCATCTACCGTCCATGAAGAGGCAATCGGCCACGCA 1038
Db 1081 CCCCMAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGATGCTGTGATCCAGCAC 1140
Qy 1039 TGAAGTGGCCGACGACAGCTGGAGCTCGCAAGAACAGCTTGGAGCTCTACGGGGCTG 1098
Db 1141 TTCAGACCTCCCAAGSACACGTCAGTGTGGAAGGCCAGCTTTGAGCTCTATGGGGCTG 1200
Qy 1099 ACTTCGCTCTGGAGGGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCCCCACCA 1158
Db 1201 ACTTCGTTTCGGGAGGAGTTCAGCCCTGCTGATTGAGATCAACGCCAGGCCCAACA 1260
Qy 1159 TGACCCCGTCCAGCGGTCAAGGCCCAAGCTGTGTGCAAGGTGACAGGAGGACACATCA 1218
Db 1261 TGGACCCCTCCACAGCAGTCACTGCCCGGCTCTGTCTGGCGTGCAAGCTGACACCCCTGC 1320
Qy 1219 AGGTGGCC-----GTGGACCGCAGCTGTGACATCGGCACTTCGAGCTCTCT 1266
Db 1321 GCGTGGTCAATTGACCGGATGCTGGACCGCAACTGTGACACAGAGGCTTTGAGCTCATCT 1380
Qy 1267 GGAGGACGCGGTGGTTGAGCCGCCCCCAATTCAGCGGGTCCGACCTCTCGCTGGCGGGCG 1326
Db 1381 ATAAAGAGCGCTGTGTGGAGGTGCTCAATATGTGGGCATCCGGCTCTCTGTGTAGGGCT 1440
Qy 1327 TCAGTGTGAGGAGAGCCAGAGGACAGGTGTGCGCCCTGTGCAACCTCAAGGCGCTCGGCT 1386
Db 1441 TCACCATCAAGAGGCCCATGGCGATGTGTATCGGCGGATGGGGTCCGCCACGAGCTCC 1500
Qy 1387 CGCTGTGGACGCGCAGCCGCTGAAGGACAGGGGCCCCCTCGGCCATGCC-AGACCTCGCC 1445
Db 1501 CTCTGCTGACCCAGCAGGCTCTTGGGGAAGGCAAGGACTCGGGGATTCCTTACCCACAGGT 1560

Qy 1446 CAGGAGACCCCATCATCAGCTCTCAGCGGGACTTGGGACTGAAAGAGAGGGGCTC 1505
Db 1561 CAGCTTCTAGGAAAGGCACTGGGGCCAGGAGCTGGGGCACAGTGAAGACCACTCTCCA 1620
Qy 1506 CCCCTGGCTTCTGTCGACCCCTTAAGGGGGGAGCCGAGAGCGGTGGAGCCGACAGCCC 1565
Db 1621 CTGCCACCACTTTCAGCCCCCGGAAAGGGAAGGCAAGGCAAGGCAAGGCAAGCCC 1680
Qy 1566 ACCCGCACCAAGCTGCTGGGAAGGTGGAGCTCCCGGC 1603
Db 1681 TGGTCTGCCCAATCTCTGGGAGTGGGATGCCCCGAGC 1718

RESULT 15

US-10-029-386-24894/c
; Sequence 24894, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Rank, David R.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GE
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24894
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022327.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.35
; OTHER INFORMATION: SWISSPROT HIT: P38160, EVALUE 5.00e-04
; OTHER INFORMATION: EST HUMAN HIT: BE82753.1, EVALUE 3.00e-06
; OTHER INFORMATION: NT HIT: gil14726145, EVALUE 8.00e-09
US-10-029-386-24894

Query Match 23.2%; Score 427; DB 16; Length 490;

Best Local Similarity 100.0%; Pred. No. 3e-107;

Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 848 CAGCGCATCCACCTGTGCAACAACGCGCTCCAGAACTACCTGAAAGATGATGTGGCCG 907
Db 427 CAGCGCATCCACCTGTGCAACAACGCGCTCCAGAACTACCTGAAAGATGATGTGGCCG 368
Qy 908 CAGCCCCCTGTGCGCCGCAACAACATGTGGACCAAGCTTCCAGGAGTACCTGCA 967
Db 367 CAGCCCCCTGTGCGCCGCAACAACATGTGGACCAAGCTTCCAGGAGTACCTGCA 308
Qy 968 GCGCAGGCGCGTGGCGCGGTGGGGAGCGCTCATCTACCCGCTCCATCAAGAGGCCAT 1027
Db 307 GCGCAGGCGCGTGGCGCGGTGGGGAGCGCTCATCTACCCGCTCCATCAAGAGGCCAT 248
Qy 1028 CCGCCACGCGCATGAAGTGGCCAGGACCACTGGAGCTCGCAAGAACAGCTTTGAGCT 1087
Db 247 CCGCCACGCGCATGAAGTGGCCAGGACCACTGGAGCTCGCAAGAACAGCTTTGAGCT 188
Qy 1088 CTACGGGGCTGACTTTCGTCTTTGGGAGGGAATTTCAGGGCCCTGGCTGATCGAGATCAATT 1147
Db 187 CTACGGGGCTGACTTTCGTCTTTGGGAGGGAATTTCAGGGCCCTGGCTGATCGAGATCAATT 128
Qy 1148 CAGCCCCCAGGAGCGCCGCTCCAGCGCGCTCAGCGCCAGCTGTGTGACAGGTGCAGGA 1207
Db 127 CAGCCCCCAGGAGCGCCGCTCCAGCGCGCTCAGCGCCAGCTGTGTGACAGGTGCAGGA 69
Qy 1208 GGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGCATCGGCAACTTCGAGCTCTGTG 1267
Db 67 GGACACCATCAAGGTGGCGGTGGACCGCAGCTGTGCATCGGCAACTTCGAGCTCTGTG 8

Qy 1268 GAGGAG 1274
Db |||||
7 GAGGAG 1

Search completed: September 24, 2005, 16:48:37
Job time : 1423.93 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 20:40:04 ; Search time 335.858 Seconds
(without alignments)
8954.609 Million cell updates/sec

Title: US-10-615-659-1
Perfect score: 1838
Sequence: 1 tggagcagcctggggcccc.....actctcccaaggcggaattc 1838

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	421.4	22.9	3001	4	US-09-620-312D-494
2	56	3.0	909	4	US-09-252-991A-5992
3	56	3.0	1305	4	US-09-252-991A-5951
4	52.6	2.9	1272	4	US-09-252-991A-2471
5	52.6	2.9	1449	4	US-09-252-991A-2200
6	52.6	2.9	2790	4	US-09-252-991A-2632
7	50.4	2.7	1524	4	US-09-807-897-3
8	50.2	2.7	9039	4	US-09-902-540-983
9	48.6	2.6	2277	1	US-08-676-967-5
10	48.6	2.6	2277	1	US-08-676-974-5
11	48.6	2.6	2277	2	US-09-098-487-5
12	48.2	2.6	2004	1	US-08-471-033-18
13	48.2	2.6	2004	2	US-08-471-044-18
14	48.2	2.6	2004	2	US-08-463-483A-18
15	48.2	2.6	2004	2	US-08-471-046A-18
16	48.2	2.6	2004	2	US-08-470-566B-18
17	48.2	2.6	2004	2	US-08-469-334-18
18	48.2	2.6	2004	3	US-09-300-529-18
19	48.2	2.6	2655	1	US-08-471-033-17
20	48.2	2.6	2655	2	US-08-471-044-17
21	48.2	2.6	2655	2	US-08-463-483A-17
22	48.2	2.6	2655	2	US-08-471-046A-17
23	48.2	2.6	2655	2	US-08-470-566B-17
24	48.2	2.6	2655	2	US-08-469-334-17
25	48.2	2.6	2655	3	US-09-300-529-17
26	48	2.6	5314	4	US-09-620-312D-138
27	47.6	2.6	927	4	US-09-902-540-9188

28	47.6	2.6	4884	4	US-09-252-991A-12126	Sequence 12126, A
29	47.6	2.6	4884	4	US-09-252-991A-12292	Sequence 12292, A
30	46.4	2.5	4600	4	US-09-702-705-1797	Sequence 1797, Ap
31	46.4	2.5	4600	4	US-09-736-457-1797	Sequence 1797, Ap
32	46.4	2.5	4600	4	US-09-671-325-1797	Sequence 1147, Ap
33	46.4	2.5	19455	4	US-09-302-540-1147	Sequence 1147, Ap
34	46	2.5	1218	4	US-09-902-540-5379	Sequence 5379, Ap
35	46	2.5	2847	4	US-09-302-540-7588	Sequence 7588, Ap
36	46	2.5	7704	4	US-09-902-540-743	Sequence 743, App
37	46	2.5	34552	4	US-09-902-540-1262	Sequence 1262, Ap
38	45.8	2.5	1515	4	US-09-302-540-9184	Sequence 9184, Ap
39	45.8	2.5	9556	4	US-09-302-540-982	Sequence 982, App
40	45.6	2.5	534	4	US-09-252-991A-7596	Sequence 7596, Ap
41	45.6	2.5	576	4	US-09-252-991A-7886	Sequence 7886, Ap
42	45.6	2.5	1047	4	US-09-252-991A-14678	Sequence 14678, A
43	45.6	2.5	1566	4	US-09-902-540-4578	Sequence 4578, Ap
44	45.6	2.5	1686	4	US-09-252-991A-14548	Sequence 14548, A
45	45.6	2.5	1695	4	US-09-252-991A-7816	Sequence 7816, Ap

ALIGNMENTS

RESULT 1
US-09-620-312D-494
; Sequence 494, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: Pt_Fl_genes Version 1.0
; SEQ ID NO 494
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1361)..(2419)
US-09-620-312D-494

Query Match 22.9%; Score 421.4; DB 4; Length 3001;
Best Local Similarity 63.8%; Pred. No. 8.4e-90;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 232 GGGAGGAGCGCGGAGCAGCGACCTCAGCAGCAGGCAAGATGCTGAAATGCTCAGGCAA 291
|||||
Db 1223 GGGATCGTGACAGCCCTCCCTATGTGCGAGGAGACAGCAGCCCAAGAACAGGAGAAA 1282
|||||


```
QY 292 AGCTCAG-GGGCCCTCCGGGGGCGAGCTTGTGGACATCGCGTGCAGAGTGTGCCAGGCCCTAC 350
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1283 AACCAGCTTGTGTGTCGCCAGAGTTTGTGGATGAAGCTCTGTGTGCGTCGCGAGGAGTAC 1342
QY 351 CTGGGCGAGTGGAGCATGAGGACATCGACGTCAGCAGATGCCGTGGAGGACCTCACT 410
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1343 CTTAGCAACTTGGCCACATGGACATCGACAAGGACCTGGAGGCCCGCGTGTACTCTACC 1402
QY 411 GAGGCCGAGTGGAGGACCTGACCCAGCAGTACTACTCTCCTCGTTTCATGGCCGATGTTTC 470
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1403 CCGAGGGCTGGTCCCTCTCTCCAGCGCTACTACCAAGTGGTCCACGAAGGGGACGAA 1462
QY 471 ATCTCAATCAAGAAATTAATTTTCGAGTGGCCAGGCTCTGCTGAATAGAAATCAAGTCT 530
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1463 CTCAGGCACCTCGACACTCAGGTTCAGCGCTGTGAGGACATCTGCGAGCAGCTGCAGGCC 1522
QY 531 GTGAACCTTCAGACGACATTTAGCGGCTCCGGACATCTGGATATTAAGCCCGCGGCC 590
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1523 GTGGTACCCAGATAGACATGAAGGGGATCGCAACATCTGGATCGTGAAGCCAGAGCC 1582
QY 591 AAGTCCCGGGCGGACACATAGTGTGCATGGACCGTGTGGAGGAGATCTCTGGAGCTGGCA 650
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1583 AAGTCCCGGGAGGACATCATGTGCATGGACCCCTGGAGGAGATGCTGAAGCTGGTG 1642
QY 651 GTGCAGACACCTCTTTCCAGGGACAAAGAGTGGTGGTCCAGAGTACATCGAGAG 710
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1643 AACGGCAACCCCGTGTGTATGAAGGACGGCAAGTGGTGGTGCAGAAATATATTGAGCG 1702
QY 711 CGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTCTCTCTCAGGACTGG 770
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1703 CCCCCTCATCTTTGGCACCAAGTTTGAACCTCAGACAGTGGTCTCTGTAACTGACTGG 1762
QY 771 AACCCCTCACCATCTGGTTTACAGGAGAGTTACTTGGCGTTCTCAACTCAGGCGTTTC 830
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1763 AACCACATTACCTGTGGTTTACCGCGACAGCTATATCCGCTTTTCCAGCAGCCCTTC 1822
QY 831 TCCTGGAACAAGTGACAGGCGCATTCACCTGTGCAACAAAGCGGTCCAGAAATACCTG 890
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1823 TCCTGGAAGAAGCTGACAACTCAGTGCACTGTGCAACAACTCCATCCAGAAAGACCTG 1882
QY 891 AAGATGATGTGGCGCAGCCCGCTGCTGCCGACACAAATGTGGACCAATGTGGTCTAGCCAGAG 950
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1883 GAGAACTCATGCGCATCGGCATCCACTGCTTCCGCCAGACAAATGTGGTCTAGCCAGAG 1942
QY 951 TTCCAGGAGTACTCTGAGCGCCAGGGCCGCTGGCGCGCTGTGGGCGAGCGTCACTACCG 1010
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1943 TTCCAGGCCCATCTGACGAGATGGTGGCCCCAAATGCTTGTTCACCATCATCTGTCCT 2002
QY 1011 TCCATGAAGAAGGCCCATCGCCCAAGCATGAAGTGGCCCGAGACCACTGAGGCTCGC 1070
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2003 GGCATGAAGGATGCTGTGATCCAGCAGCTTCAGACCTCCAGGACACACCGTGCAGTGTGG 2062
QY 1071 AAGAACAGCTTTGAGCTCTACGGGCTGACTTCTGCTTGGGAGGACTTCAGGCCCTGG 1130
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2063 AAGGCAGCTTTGAGCTCTATGGCGCTGACTTCTGCTTTCGGGAGGAGACTTCAGGCCCTGG 2122
QY 1131 CTGATCGAGATCAATTCAGGCCCCACCATGCAACCGCTCCACGCGGTCAACGCCCGCAGCTG 1190
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2123 CTGATTTGAGATCAAGCCAGCCCCAGATGGCACCTCCACAGCAGTCACTGCCCCGGCTC 2182
QY 1191 TGTGCACAGGTGAGGAGACACCATCAAGGTGGCC-----GTGACCGCAGC 1238
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2183 TGTGCTGGCGTCAAGCTGACACCTTGCCTGCGCTGGTCAATTGACCGGATGTGAGCCGCAAC 2242
QY 1239 TGTGACNTCGCAACTTTCAGCTCTGTGGAGGACGCGGT 1279
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2243 TGTGACACAGGACCTTTGAGCTCATCTATAAGCAGCCCGT 2283
```

RESULT 2

US-09-252-991A-5992

; Sequence 5992, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5992

; LENGTH: 909

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5992

Query Match

Best Local Similarity 3.0%; Score 56; DB 4; Length 909;

Matches 224; Conservative 44.4%; Pred. No. 0.0013;

Matches 224; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

QY 932 CATGTGGACACGAGTTCAGAGTACCTGCAGCGCCAGGCGCGTGGCGCGTGG 991

Db 25 CCTGATGAGCTTCACCGGGCGCGCAACAGGAAGCGCGGTCCGGTGAAGTCCG 84

QY 992 GGGCAGCGTCATCTACCGCTCATGAAGAAGGCCATCGCCACGCGCATGAAGGTGGCCCA 1051

Db 85 CGTGGCGTGCAGGATATCTTCAACGGGCTGTACTCTCCACCGCGGTGCTCGCGCCCT 144

QY 1052 GGAACAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTGACTTGGTCTTGG 1111

Db 145 GCGCCACCGCGACGTGAGCGCATCGGCAGCACATCGACATGGCCCTGCTCGAGTGA 204

QY 1112 GAGGACATTCAGGCGCTGGCTGATCGAGATCAATTCAGCGCCCAACATGACCCGCTCCAC 1171

Db 205 GGTGCGCTGCTGGCGCAACAGACCTTCAACTACTGACCCCGCGGTGCGCGCGGG 264

QY 1172 GCGCGTCAAGCGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGGCGGTGA 1231

Db 265 CTGGGCAACGACACACCGCGACATCGTCCGTTACCGACTTCCCGACCGCGAGCGGCA 324

QY 1232 CGCAGCTGTGACATTCGGCAACTTCGAGCTCTGTGGAGGCGAGCGGTGTTGAGCGCGC 1291

Db 325 CATGATCTCTACGCTCGGCAACGACAGCAGTTCCGCAAGTTTCGCGAGCTGCGCAGACA 384

QY 1292 CCCATTACGGGTCCGACCTTCGCTGGCGGGCTGAGTGTGAGGAGAGCCAGGAGCA 1351

Db 385 CCGGAAATGGGCGGACGACCGCGCTTCGCCAACCAAGGCGCGGTGGCCAAACCGGCA 444

QY 1352 GGTGCTGCGCGCTCTCAACCTCAAGGCTTCGGCTTCGCTGTTGGACGCGCGAGCGGTGAA 1411

Db 445 GGTGCTGATCCGCTGATCCGCGAGCCACCGTGTGTCACACCCGCGAGTGTGATTCT 504

QY 1412 GGCACGGGCGCTTCGGCGATGCC 1435

Db 505 TTCCCTGGAGCGCGCGCGTGCC 528

RESULT 3

US-09-252-991A-5951

; Sequence 5951, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5951
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5951

Query Match 3.0%; Score 56; DB 4; Length 1305;

Best Local Similarity 44.4%; Pred. No. 0.0014;
Matches 224; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

Qy	932	CATGTGACACAGACACAGGTTCAGGAGTACCTGCAGCGCAGCGCGTGGCGCGTGGT	991
Db	546	CCTGATGAGCTACCGGGCGCGGCAACAGAGAGGCGCGTCCGGTGAAGTGG	605
Qy	992	GGGCGAGCTCATCTACCGCTTCATGAAGAGGCCATCGCCACGCCCATGAAGGTGGCCCA	1051
Db	606	CGTGGCGTGACGATATCTCACCGGCTGTACTCTCCACCGGCTGCTCGCGGCCCT	665
Qy	1052	GGACCGTGGAGCTCGAAGAACAGCTTTGAGCTTACGGGGCTGACTTGTCTCTGG	1111
Db	666	CGCCCAACCGGAGCTCAGCGGCGATCGGCCAGCACATCGACATGGCCCTGCTCGACGTGA	725
Qy	1112	GAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGCGCCCAACCATGCACCGTCCAC	1171
Db	726	GGTGGCTGTGCTGGCCCAACAGACCTCTCACTACCTGACCAACCGGCTGCGCGCGGCG	785
Qy	1172	GCGGCTCACGGCCAGCTGTGTGCACAGGTGCAGGAGGACACCATCAAGGTGGCGTGA	1231
Db	786	CCTGGGACGACACCCGAAACATCGTCCGTACCAAGTCTTCCCAACCGCCAGCGGA	845
Qy	1232	CCGAGCTGTGACATCGCAACTTCGAGCTCTGTGTGAGGAGCGCGTGTGTGAGCGCC	1291
Db	846	CATGATCTCAGCTGCGCAACGACAGCAGTTCGCAAGTTCGCGGAGCTGCGCAGACCA	905
Qy	1292	CCCATTCAGCGGTCGACCTCTCGTGGCGGGCTCAGTGTGAGGAGACCGAGGGA	1351
Db	906	CCCGGAATGGCGAGACCGCGCTTCGCCCAACAAAGCGCGGTGGCCCAACCGCGA	965
Qy	1352	GGTGTGCGCGCTGTGCAACCTCAAGGCTCGGCTCTGTGAGCGCGAGCGCTGMA	1411
Db	966	GGTGTGATCCCGTGTATCGCCAGGCGACGCTGTGACACCAACCGCGCGATCTCT	1025
Qy	1412	GGCAGGGGCGCTTCGGCCATGCC	1435
Db	1026	TTCCCTGGAGGCGCGCGCTGCC	1049

RESULT 4

US-09-252-991A-2471
; Sequence 2471, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2471
; LENGTH: 1272

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2471

Query Match 2.9%; Score 52.6; DB 4; Length 1272;
Best Local Similarity 48.5%; Pred. No. 0.009;

Matches	145;	Conservative	0;	Mismatches	154;	Indels	0;	Gaps	0;
Qy	545	GGACATTGACGGCTCGGAAACATCTGGATTATAAGCCCGCCAGTCCCGGGCCG	604						
Db	756	GGAAATGCTCGCGACGGCGAGGTACGACGATTTCCGGCACCGGGAGGACGATCCT	815						
Qy	605	AGACATAGTGTGCATCGACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCA	664						
Db	816	CGACGTACGATGGTTCGACCGCGGCGCGCTGGAGGTGTCGAGTAGGCGCAGGC	875						
Qy	665	TCCTTTCCAGGGACAAAGTGGTGTCCAGAAAGTACATCGACAGCCCGCTGTCTG	724						
Db	876	GGTACGCTGGAAGTTCGAGGTTCGCCAGGACATCGAGCGGTGATCTCGGCTT	935						
Qy	725	TGACACCAAGTTGCATCAGACAGTGGTTCCTCGTACGAGTGGACACCCCTGACCAT	784						
Db	936	CATGATCAAGGACCGCTCGGCCAGCGATGTACGGGATCAATACCCACCGCTGGCAA	995						
Qy	785	CTGGTTCTCAAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC	843						
Db	996	GGCGCTCAGGACCTGAAGCGCGGAGCGCATCACTACCGCTTGGCTTCGACATGC	1054						

RESULT 5

US-09-252-991A-2200/c
; Sequence 2200, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2200
; LENGTH: 1449

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2200

Query Match 2.9%; Score 52.6; DB 4; Length 1449;
Best Local Similarity 48.5%; Pred. No. 0.0094;

Matches	145;	Conservative	0;	Mismatches	154;	Indels	0;	Gaps	0;
Qy	545	GGACATTGACGGCTCGGAAACATCTGGATTATAAGCCCGCCAGTCCCGGGCCG	604						
Db	562	GGAAATGCTCGCGACGGCGAGGTACGACGATTTCCGGCACCGGGAGGACGATCCT	503						
Qy	605	AGACATAGTGTGCATCGACCGTGTGGAGGAGATCTTGGAGCTGGCAGCTGCAGACCA	664						
Db	502	CGACGTACGATGGTTCGACCGCGGCGCGCTGGAGGTGTCGAGTAGGCGCAGGC	443						
Qy	665	TCCTTTCCAGGGACAAAGTGGTGTCCAGAAAGTACATCGACAGCCCGCTGTCTG	724						
Db	442	GGTACGCTGGAAGTTCGAGGTTCGCCAGGACATCGACGGCTGATCTCGGCTT	383						
Qy	725	TGACACCAAGTTGCATCAGACAGTGGTTCCTCGTACGAGTGGACACCCCTGACCAT	784						
Db	382	CATGATCAAGGACCGCTCGGCCAGCGATGTACGGGATCAATACCCACCGCTGGCAA	323						
Qy	785	CTGGTTCTCAAGGAGAGTTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAAGC	843						
Db	322	GGCGCTCAGGACCTGAAGCGCGGAGCGCATCACTACCGCTTGGCTTCGACATGC	264						

RESULT 6

US-09-252-991A-2632
; Sequence 2632, Application US/09252991A


```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2632
; LENGTH: 2790
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2632

Query Match      2.9%; Score 52.6; DB 4; Length 2790;
Best Local Similarity 48.5%; Pred. No. 0.012;
Matches 145; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 545 GCACATTGACGGCTCCGGAACATCTGGATTATAAGCCGCGGCAAGTCCGGGGCG 604
Db 895 GGAATGCTCGCGGACCGGACGATACGACGATTTCCGCGACCGGCGAGCGCATCCT 954
QY 605 AGACATAGTGTGCATGGACCGGTGTGGAGGAGATCTCTGGAGCTGGGAGCTGCAGACCCACC 664
Db 955 CGACGTACGATGGTTCGACCGGCGAGCGCGCTGGAGGTGTGAGTAGGCGAGC 1014
QY 665 TCTTTCCAGGGACAAAGTGGTGTCTCCAGAGTACATCGAGACCGCGCTCTCATCTG 724
Db 1015 GGTGACGCTGGAAGTCGAGGTCGAGGTTCCGCCAGGACATCGAGCGCTCATCTCGGCTT 1074
QY 725 TGACACCAAGTTCGACATCAGACAGTGGTTCCTGTCAGGAGTGAACCCCTGACCAT 784
Db 1075 CATGATCAAGGACCGCTCGCGACCGGATGTACGGGATCAATACCACCGCTTGACAA 1134
QY 785 CTGGTTCTAAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTCGACAAAGC 843
Db 1135 GCGCTCACCAGCCTGAAGCGCGGAGCGCATCCTACCGCTTGGCTTGCATGTC 1193

RESULT 7
US-09-807-897-3
; Sequence 3, Application US/09807897
; Patent No. 6660851
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Yoshimitsu
; APPLICANT: UEKI, Jun
; TITLE OF INVENTION: NOVEL DNA FRAGMENT ENHANCING GENE EXPRESSION LEVEL
; FILE REFERENCE: 0230-0156P
; CURRENT APPLICATION NUMBER: US/09/807,897
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: PCT/JP00/05539
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Oryza sativa L.
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Variety: IR24 - Tissue: pistil
; OTHER INFORMATION: Library name: lambdaZAPII cDNA library derived from pistil mRNA
; OTHER INFORMATION: Clone name: RPC213
; NAME/KEY: misc feature
; OTHER INFORMATION: Strandedness: double - Topology: linear
; OTHER INFORMATION: Molecular type: cDNA to mRNA
; OTHER INFORMATION: Feature: nt1, nt2, nt3; transcription initiation points of RPC213
; OTHER INFORMATION: gene determined by primer extension
```

```
; NAME/KEY: misc feature
; OTHER INFORMATION: nt22-nt24: the first ATG of RPC213 gene
; OTHER INFORMATION: nt134-nt136: the second ATG of RPC213 gene
; OTHER INFORMATION: nt295-nt297: the third ATG of RPC213 gene
; OTHER INFORMATION: nt1276-nt1278: termination codon of RPC213 gene
; NAME/KEY: misc feature
; OTHER INFORMATION: nt1343-nt1348: nt1365-nt1370: poly (A) signals
; OTHER INFORMATION: nt1507-nt1524: poly (A)
US-09-807-897-3

Query Match      2.7%; Score 50.4; DB 4; Length 1524;
Best Local Similarity 47.2%; Pred. No. 0.032;
Matches 187; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 963 CTGCAGCGCCAGGGCGGTGGCGCGCTGCGGGGACGCTCATCTACCGCTCCATGAAGAAG 1022
Db 796 CAGCAGCGGCGACCGGGCGCTGCCAAGATCTCGGGGTGATCGCGCGACGAGGCGG 855
QY 1023 GCCATCGCCACGCCATGAAGGTGGCCAGGACACGAGGAGCTTCGCGAAGAACAGCTTT 1082
Db 856 CACGAGGGCGGGCTACACGATGGCTCCGCGAGGCTGTTTCGAGCTCGACCCGAGCGCATG 915
QY 1083 GAGCTCTACGGGGCTGACTTCGCTCTTGGGAGGACTTCAGGCCCTTGGCTGATCGAGATC 1142
Db 916 GCGCGCGGCTGCGCGAGCTCATGCGGGGAAGGTGACCATCGCGGGGAGCTCATGTGCG 975
QY 1143 AATTCAGCCCCACCATGCACTCCGCTCCAGCGCGTCAAGCGCCAGCTGTGTGCACAGGTG 1202
Db 976 GAGCGCGGCGGCGGCGGCGAGCACAGCTGTTC---GCCCGGTCTCCGCGGTGGCG 1032
QY 1203 CAGGAGGACACCATCAAGGTGGCCCTGGACCGGAGCTGTGATCGGCAACTTCGAGCTC 1262
Db 1033 GAGCGCGCGGCTGTACACGCGGAGGACTACGCGGAACTCGTCGAGCACTTCGTGCGG 1092
QY 1263 CTGTGAGGCGACCGGTGTTGAGCGCGCCCTCATTCAGGGGTTCGACTCTCGGTGGCG 1322
Db 1093 AGTGGCGGGTGGCGGAGCTCGCGCGGGGCTCTCCGCGAGGCGGCGCGGAGGAG 1152
QY 1323 GCGGTGAGTGAGGAGAGCCAGGAGGAGGAGGTGCTG 1358
Db 1153 TACCTGTGGGGTTGGCGCCCAAGATCCGAGGATG 1188

RESULT 8
US-09-902-540-983
; Sequence 983, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 983
; LENGTH: 9039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-983

Query Match      2.7%; Score 50.2; DB 4; Length 9039;
Best Local Similarity 46.9%; Pred. No. 0.065;
Matches 190; Conservative 0; Mismatches 213; Indels 2; Gaps 1;

QY 1241 TGACATCGCAACTTCGAGCTCTCTGTGGAGGAGCGGCTGTGAGCGCGCCCTTCAG 1300
Db 2988 TGACCTCGCTCTCGATATATCCCGAGGAGCTGCCAAGAGCGCTCCCGCGGTGAAGCA 3047
```


QY 1301 CGGGTCGACCTCTGCGTGGCGGCGTCACTGTGAGGAGAGCAGGAGCGGTGCTGCC 1360
Db 3048 AGCGTGTGCTGTGCTGCGGCTCTACTGACCGCTTGAAGGAAGAGCCCGTGGG 3107
QY 1361 CGTCTGCAACC--TCAAGGCTCGGCTCGCTGTGAGACGCGCAGCGCTGAAGGCACGG 1418
Db 3108 AAGATGAGGCCATTCCTCCGCGCGGGGATGAAGGACGCTCCGCTCCCGCGGAAGCC 3167
QY 1419 GGGCCCTCGGCGCATGCCAGACCTTGCACGAGGACCCCATCACAGCTTCCAGCGGGAC 1478
Db 3168 GCTCCCGCGGCTCGCGGAGCACCCTGAGGACCCCTCAAAGGTGCTCTCCGCTGGAAG 3227
QY 1479 TTGGGACTGAAGGAAGAGAGGCGCTCCCTCTGCGCTTGTGCGACCTTTAAGGGGGCA 1538
Db 3228 ATTGCCAGGGCGGCGCTCTGTATCCGCTTGAATCAGGCGCACCGCGCGCGGCC 3287
QY 1539 GCGGAGCGGTGGAGCGGACACCGCACCCCGCCCAAGCTGCTGGGAAGGTGGAGCTC 1598
Db 3288 GCCAGCAGCGAGGAAGCGCGCGCCCCCAAGAGCGCCCAAGGGCAAGGCAAGGAGGCC 3347
QY 1599 CGGGCTGCGCTGTGCGCAGTGGACAGTGCAGTCAGCGGCCCAAGACCC 1643
Db 3348 GCGCCGCTCGCGCGCGCGCGCGCTCTGCGCGGCGCGCTCACC 3392

RESULT 9

US-08-676-967-5
; Sequence 5, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-967-5

Query Match 2.6%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.097;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 833 CTTGGACAAGCTGGACAGCGCCATCCACCTGTGTGCAACAGCGCGTCCAGAAGTACCTGAA 892
Db 1437 CCAGAACATCTTCTGAGCGCGCACCGCTGTGCTGCGACAACTGCGCCAGGCCGTGGA 1496

QY 893 GAATGATGTGGCGCGCAGCCCCCTGCTGCCGCACACAAATGTGGACCAAGCAGGTT 952
Db 1497 CGACAAGAGCTCGCAAGCTGCTGTGAGCGCCACAGCGGAGAGGGCGTGCAT 1556
QY 953 CCAGGAGTACCTGTCAGCGCGCCGTCGTCGCGCGCTGTGCGGCGAGCGTCTATCTACCGTC 1012
Db 1557 CAAGGAGTCCGCGTGTATGCGGACCTGAAGGCGTGCACGCGCAATGAAGGGCCAGAG 1616
QY 1013 CATGAAGAGGCCATCGCCACCGCATGAAGTGGCCCGAGGACCACTGTGAGCGCTCGCAA 1072
Db 1617 CTTGGGCTAGCGCTTCGCGAGTTCCAGGAGCAGGACGCGCTGAAGGCGCTGCGCCT 1676
QY 1073 GAACAGC 1079
Db 1677 GATCAAC 1683

RESULT 10

US-08-676-974-5
; Sequence 5, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-974-5

Query Match 2.6%; Score 48.6; DB 1; Length 2277;
Best Local Similarity 49.8%; Pred. No. 0.097;
Matches 123; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 833 CTTGGACAAGCTGGACAGCGCCATCCACCTGTGTGCAACAGCGCGTCCAGAAGTACCTGAA 892
Db 1437 CCAGAACATCTTCTGAGCGCGCACCGCTGTGCTGCGACAACTGCGCCAGGCCGTGGA 1496

QY 893 GAATGATGTGGCGCGCAGCCCCCTGCTGCCGCACACAAATGTGGACCAAGCAGGTT 952

Db 1497 CGACAAGAGCTCGCAAGCTGCTGTGAGCGCCACAGCGGAGAGGGCGTGCAT 1556

QY 953 CCAGGAGTACCTGTCAGCGCGCGCGCTGCGCGCTGTGCGGCGAGCGTCTATCTACCGTC 1012

Db 1557 CAAGGAGTCCGCGTGTATGCGGACCTGAAGGCGTGCACGCGCAATGAAGGGCCAGAG 1616

Qy	1013	CATGAAGAGGCCATCGCCACGCCCATGAGTTGGCCACGACACAGTGGAGCCTCGCA	1072
Db	1617	CTTGGGCTAGCCCTTCGCCAGTTCAGGAGCACGAGACGCGCTGAAGGCCCTGGCGCT	1676
Qy	1073	GAAACAGC	1079
Db	1677	GATCAAC	1683

```

RESULT 11
US-09-098-487-5
; Sequence 5, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESS: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Osman Ph. D., Richard A
/ REGISTRATION NUMBER: 36,627
/ REFERENCE/DOCKET NUMBER: UCB96-055
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415)343-4341
/ TELEFAX: (415)343-4342
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2277 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ US-09-098-487-5

```

	Query Match	2.6%;	Score 48.6;	DB 2;	Length 2277;
	Best Local Similarity	49.8%;	Pred. No. 0.097;		
	Matches 123;	Conservative	0;	Mismatches 124;	Indels 0; Gaps 0;
QY	833	CTTGGAACAAGCTGAGACGCCCATCCACCTGTGCAACAAGCCGTGCCAGAACTCTGAA	892		
DB	1437	CGAAGACATCTTCTGTGAGCCGACCCGCTGTGCTGTGCACAACTGTGCCAAGGCCGTGGA	1496		
QY	893	GAATGATGTGGGCGCGAGCCCTCTGTGTCGCCGACACAAATGTGGACACGACACAGGTT	952		
DB	1497	CGACAAGCAGCTCGCAAGCTGTGCTGTAGCGGCCACGAGCGGCGAAGGCGTGGGCAT	1556		
QY	953	CGAGGAGTACCTCAGCGCCAGGCGCGTGCGCCGTGTGGGGCAGGCTCATCTACCGTC	1012		
DB	1557	CAAGGAGTCCGGGTGATGCGCACTGAAGGGGTGCACGGCAACATGAAGGGCCAG	1616		
QY	1013	CATGAAGAAGGCCATCGCCACGCCATGAAGGTGGCCGAGGACCACTGGAGCCTCGCA	1072		
DB	1617	CCTGGGCTACGCTTCTGCCGAGTTCCAGGAGCAGCAGACGCCCTGAAGCCCTGGCCCT	1676		

```

RESULT 12
US-08-471-033-18
; Sequence 18, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB7
US-08-471-033-18

```

	Query Match	2.6%	Score 48.2;	DB 1;	Length 2004;
	Best Local Similarity	43.7%	Pred. No. 0.12;		
	Matches 212;	Conservative	0;	Mismatches 273;	Indels 0; Gaps 0;
Qy	659	CCACCCCTCTTTTCAGGGGACAAACAGTGGGTGGTTCAGAAGTACATCAGAGCGCGCGTGCT	718		
Db	720	CCACCCCATCACCTGTGAACAAAGAGCAGGTGTGGACAACTGCTGAACAAACAGGCCCATGAT	779		

QY 719 CATCTGTGACCAAGTTTCAGACATCAGACAGTGGTTCTCTGTCACGACTGGAAACCCCT 778
Db 780 GCTGGAGACCAACAGACCGCGCTCTACAGATCAAGACACCCACGCGACATCGT 839
QY 779 GACCATCTGTTTCAAGAGAGTTACTTGGCGTTTCTCAACTCAGCGCTTCTCCCTGGA 838
Db 840 GACCGCGCGAGTGGAAACGCGGTGATCCAGCAGATCAAGGCCAAGACCCGACGATCAT 899
QY 839 CAAGCTGCAGCGCCATCCACCTGTGCAACAGCGCTCCAGAGTACCTGAGATGA 898
Db 900 CTGCGACGACGCGCGCTGCGCCGAGAGCGGTGGCCGCGCCCAAGGACTACGAGAACCC 959
QY 899 TGTGGCGCGACGCGCTGCTGCGCCGACACCAACATGTGGACACGACACCGAGTTCCAGGA 958
Db 960 CGAGGACAAACCCCAAGCTGACCTGTAGAGACGCGCTGAGCTGAGCTACCCGACGA 1019
QY 959 GTACTCTGACGCGCAGCGCGCTGCGCGCGTGTGGGCGAGCGTCACTACCCGTCATGAA 1018
Db 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCT 1079
QY 1019 GAAGGCCATCGCCACGCGCATGAAGTGGCGCCGAGGACCAAGTGGAGCGCTCGCAAGACAG 1078
Db 1080 GATGACCTATCTAGACGAGAACCCGCAAGAGGAGTGAACCAAGCAGCTGAACGACACAC 1139
QY 1079 CTTTGAGCTCTACGGGCTGACTTGTCTTGTGGAGGAGCTTCAGGCGCTTGGTATCGA 1138
Db 1140 CGGCAAGTTCAGGACGTGAGCCACCTGTACGAGCTGAGCTGACCCCAAGATGAAGT 1199
QY 1139 GATCA 1143
Db 1200 GACCA 1204

RESULT 13

US-08-471-044-18
; Sequence 18, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA: US/08/471,044
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018

; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA: US 08/037,057
; APPLICATION NUMBER: 25-MAR-1993
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLv3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
; US-08-471-044-18

Query Match 2.6%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.12;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 659 CCACCTCTTCCAGGAGCAACAAGTGGTGGTCCAGAGTACATCGAGAGCGCGCTGCT 718
Db 720 CCACCTCTTCCAGGAGCAACAAGTGGTGGTCCAGAGTACATCGAGAGCGCGCTGCT 779
QY 719 CATCTGTGACCAACCAAGTTTCAGACATCAGACAGTGGTCTCTGTCACGAGCTGGAAACCCCT 778
Db 780 GCTGGAGACCAACCAAGTGGTGGTCCAGAGTACATCGAGAGCGCGCTGCTGAGTCAAGTCAAGGACACCCACGCGACATCGT 839
QY 779 GACCATCTGTTTCAAGAGAGAGTTACTTGGCGTTTCTCAACTCAGCGCTTCTCCCTGGA 838
Db 840 GACCGCGCGAGTGGAAACGCGGTGATCCAGCAGATCAAGGCCAAGACCCGACGATCAT 899
QY 839 CAAGCTGCAGCGCCATCCACCTGTGCAACAGCGGTCCAGAGTACCTGAGATGA 898
Db 900 CTGCGACGACGCGCGCTGCGCCGAGAGCGGTGGCCGCGCCCAAGGACTACGAGAACCC 959
QY 899 TGTGGCGCGACGCGCTGCTGCGCCGACACCAACATGTGGACACGACACCGAGTTCCAGGA 958
Db 960 CGAGGACAAACCCCAAGCTGACCTGAGAGAGCGCTGAGTGAAGTACCTGAGTGAAGTACCCGACGA 1019
QY 959 GTACCTGCAGCGCGAGGCGCGTGGCGCGTGTGGGCGAGCTCATCTACCCGTCATGAA 1018
Db 1020 GATCAAGGAGATCGAGGCGCTGCTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCT 1079
QY 1019 GAAGGCCATCGCCACGCGCATGAGTGGCGCCGAGGACCAAGTGGAGCGCTCGCAAGAACAG 1078
Db 1080 GATGACCTATCTAGACGAGAACCAAGGAGGTTGACCAAGCAGCTGAACGACACCC 1139
QY 1079 CTTTGAGCTCTACGGGCTGACTTGTCTTGTGGAGGAGCTTCAGGCGCTTGGTATCGA 1138
Db 1140 CGGCAAGTTCAGGACGTGAGCCACCTGTACGAGCTGAGCTGACCCCAAGATGAAGT 1199
QY 1139 GATCA 1143
Db 1200 GACCA 1204

RESULT 14

US-08-463-483A-18
; Sequence 18, Application US/08463483A
; Patent No. 5849870
; GENERAL INFORMATION:

840	840	840	GACCGGGCGAGTGGAA	CGGCGTGATCCAGCAGATCAAGGCCAAGCCCGCAGCATCAT	899
839	839	839	CAAGCTGGACAGCGCCATCC	ACCTGTGTGCAACAACCGCGTCCAGAACTGCTGAAGAATGA	898
900	900	900	CGTGCAGCA	CGGCGGCGGTGGCCGAGAAGCGCGTGGCCGCAAGGACTACGAGAACC	959
899	899	899	TGTGGCGCGAGCGCCCTGCT	GCTGCTGCCCGCAACAACATGTGGACACGACACAGGTTCCAGGA	958
960	960	960	CGAGGACAAGAC	CCCCCAGCCTGACCCCTGAAGGACGCGCTGAAGCTGAGCTTACCCCGACGA	1019
959	959	959	GTAACCTGCAGCGCCAGCGCG	CGTGGCGCGTGTGGGCGAGCGTCTATCTACCCGTCCTATGAA	1018
1020	1020	1020	GATCAAGAGAT	CGAGGGCTCTGCTGTACTACAAGAACAAAGCCCATCTACGAGAGCAGCGT	1079
1019	1019	1019	GAGGCGCATCGCCACGCGCCAT	GAACTGAAGTGGCCAGGACACGCTGGAGCGCTCGCAAGAACAG	1078
1080	1080	1080	GATGACCTATCTAGACGAG	ACAACCCCAAGGAGGTGACCAAGCAGCTGAACGACACCCAC	1139
1079	1079	1079	CTTTGAGCTCTACGGGGCTG	ACTTCTCTCTTGGGAGGAGACTTCAGGCGCTTGGCTGATCGA	1138
1140	1140	1140	CGCAAGTTCAAGGAGCTG	AGCCACCTGTACGACGTTGAAGCTGACCCCAAGATGAACGT	1199
1139	1139	1139	GATCA	1143	
1200	1200	1200	GACCA	1204	
RESULT 15					
US-08-471-046A-18					
; Sequence 18, Application US/08471046A					
; Patent No. 5866326					
; GENERAL INFORMATION:					
; APPLICANT: Warren, Gregory W					
; APPLICANT: Koziel, Michael G					
; APPLICANT: Mullins, Martha A					
; APPLICANT: Carr, Brian					
; APPLICANT: Desai, Nalini M					
; APPLICANT: Kostichka, N. Kristy					
; APPLICANT: Duck, Nicholas B					
; APPLICANT: Estruch, Juan J					
; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal					
; TITLE OF INVENTION: Protein Genes					
; NUMBER OF SEQUENCES: 50					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: No. 5866326artis Corporation					
; STREET: 3054 Cornwallis Road					
; CITY: Research Triangle Park					
; STATE: NC					
; COUNTRY: USA					
; ZIP: 27709					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Floppy disk					
; COMPUTER: IBM PC compatible					
; OPERATING SYSTEM: PC-DOS/MS-DOS					
; SOFTWARE: PatentIn Release #1.0, Version #1.30B					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/08/471,046A					
; FILING DATE: 06-JUN-1995					
; CLASSIFICATION: 435					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: US 08/463,483					
; FILING DATE: 05-JUN-1995					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: US 08/314,594					
; FILING DATE: 09-SEP-1994					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: US 08/218,018					
; FILING DATE: 23-MAR-1994					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: US 08/037,057					
; FILING DATE: 25-MAR-1993					
; ATTORNEY/AGENT INFORMATION:					

```

/ FILING DATE: 08-JUN-1993
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/463,483
/ FILING DATE: 05-JUN-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/314,594
/ FILING DATE: 09-SEP-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/218,018
/ FILING DATE: 23-MAR-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/037,057
/ FILING DATE: 25-MAR-1993
/ ATTORNEY/AGENT INFORMATION:

```



```
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CCI695/CIP3/DIV8 - SQLv4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-046A-18

Query Match      2.6%; Score 48.2; DB 2; Length 2004;
Best Local Similarity 43.7%; Pred. No. 0.12;
Matches 212; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 659 CACACCTCTTTCCAGGAGCAACAAGTGGGTGTCAGAAATACATCGAGACGCGCGTGTCT 718
Db      |||||
QY 720 CCACCCCATCACCTGAAAGAACAGAGGTGGACAACTCTGTAACAACAAGCCCATGAT 779
Db      |||||
QY 719 CATCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAAACCCCT 778
Db      |||||
QY 780 GCTGGAGACCAACAGACCCGAGCGGTCTAAGATCAAGGACACCCACGGCAACATCGT 839
Db      |||||
QY 779 GACCATCTGCTTTCTACAAGGAGAGTTACTTTCGGTCTCTCAACTCAGCGCTTCTCCCTGGA 838
Db      |||||
QY 840 GACCGGCGGAGTGGAAACGGGTGATCCAGCAGATCAAGGCCAAGACCCGACGATCAT 899
Db      |||||
QY 839 CAAGCTGGACAGCGCCATCCACCTGTGCAACAACGCCGTCCAGAAGTACCTGAAGAATGA 898
Db      |||||
QY 900 CGTCGACGACGGCGAGCGCGTGGCCGAGAGCGGTGGCCGCAAGGACTACGAGAACCC 959
Db      |||||
QY 899 TGTGGCGCGCAGCCCTCTGCTGCCGACACAACATGTGGACCAAGCTTCCAGGA 958
Db      |||||
QY 960 CGAGGACAAGACCCCTGACCTGAGAGGAGCGCTTGAAGCTGAGCTACCCCGACGA 1019
Db      |||||
QY 959 GTACCTGCAGCGCAGGCGCGTGGCGCGTGTGGGCGAGCGTCTATCCCGTCCATGAA 1018
Db      |||||
QY 1020 GATCAAGGAGATCGAGGCGCTGTGTACTACAGAACAAGCCCATCTACGAGAGCAGCGT 1079
Db      |||||
QY 1019 GAAGGCCATCGCCACGCCATGAAGGTGGCCCAAGGACCAAGTGGAGCTTCGCAAGAACAG 1078
Db      |||||
QY 1080 GATGACCTATCTAGACGAGAACACCGCAAGGAGGTGACCAAGCAGCTGAACGACACCCAC 1139
Db      |||||
QY 1079 CTTTGAGCTTACGGGCTGACTTCGTCTTGGAGGGACTTCAGGCCCTGGGTGATCGA 1138
Db      |||||
QY 1140 CGGCAAGTTCAAGGACGTGAGCCACCTGTACGACGTGAAGCTGACCCCCAAGATGAACGT 1199
Db      |||||
QY 1139 GATCA 1143
Db      |||||
QY 1200 GACCA 1204
```

Search completed: September 24, 2005, 11:30:34
Job time : 337.858 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 24, 2005, 11:08:28 ; Search time 7063.48 Seconds
(without alignments)
9904.766 Million cell updates/sec

Title: US-10-615-659-1
Perfect score: 1838
Sequence: 1 tggggagcagctggggcccc.....actctcccaaggcggaattc 1838

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1196	65.1	1259	9 AY415398	Homo sapi
2	1145.6	62.3	1257	9 AY415399	Pan trogl
3	818.8	44.5	2969	3 AK030151	Mus muscu
4	812.6	44.2	2720	3 AK029745	Mus muscu
5	768	41.8	770	7 CR745100	CR745100
6	767	41.7	1255	9 AY415400	Mus muscu
7	443.6	24.1	775	5 BU937372	AGENCOURT
8	421.4	22.9	4184	3 HSM808322	AGENCOURT
9	412.4	22.4	777	7 CK597795	Homo sapi
10	399	21.7	2520	3 AK080321	Mus muscu
11	378.4	20.6	1023	4 BM808637	AGENCOURT
12	360	19.6	1022	4 BM808516	AGENCOURT
13	321	17.5	693	4 BM717061	UI-E-EJO-
14	316	17.2	861	7 CK653583	AGENCOURT
15	307.6	16.7	896	6 CD515431	AGENCOURT
16	298.4	16.2	762	6 CD627148	56047432J
17	284.2	15.5	503	7 CK604006	AGENCOURT
18	258.8	14.1	583	7 CN349423	170004243
19	258.2	14.0	680	6 CB854808	UI-CR-ECL
20	252.8	13.8	514	7 CR746623	CR746623
21	245.4	13.4	320	1 AA144756	mr73404.r
22	231.4	12.6	406	1 AI026623	ov97b08.x
23	222.4	12.1	429	7 W31827	zb94f04.r1
24	219.2	11.9	880	5 BX448465	BX448465

25	213.2	11.6	616	7	CN349424	CN349424
26	205	11.2	453	7	W31397	W31397 zb91a01.r1
27	200.8	10.9	965	5	BX431541	BX431541 BX431541
28	200.6	10.9	580	5	BP305902	BP305902 BP305902
29	199	10.8	590	1	AL045121	DKF2p434M
30	195.6	10.6	484	2	BF830679	RC3-HT023
31	194.2	10.6	3328	3	BC028169	Homo sapi
32	192.2	10.5	414	7	CN096245	EC2CA11A
33	191.6	10.4	724	4	BG916644	62814292
34	191.4	10.4	875	5	BQ718874	AGENCOURT
35	190.4	10.4	651	7	CF999072	AGENCOURT
36	188	10.2	504	5	BQ327883	MR4-RT004
37	183.6	10.0	608	7	CN349425	170005999
38	176.6	9.6	498	1	AA551103	nk75c11.s
39	173.6	9.4	488	9	CE760083	t1gr-gss-
40	172.4	9.4	482	2	BE102364	UI-R-BT1-
41	167.6	9.1	812	6	CD648578	AUF_103_C
42	161.8	8.8	977	5	BX431542	BX431542
43	158.8	8.6	630	2	BE865183	UI-M-BH1-
44	157.2	8.6	697	7	CK961833	4076568_B
45	156	8.5	850	5	BX849970	BX849970

ALIGNMENTS

RESULT 1
AY415398
LOCUS
DEFINITION Homo sapiens HCM544 gene, VIRTUAL TRANSCRIPT, partial sequence,
AY415398 1259 bp DNA linear GSS 12-DEC-2003
ACCESSION AY415398
VERSION AY415398.1 GI:39771357
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1259)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,P., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1259)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,P., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.

FEATURES
source
1..1259
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1259
/locus_tag="HCM5544"

ORIGIN
Query Match 65.1%; Score 1196; DB 9; Length 1259;
Best Local Similarity 100.0%; Pred. No. 2.2e-252;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 136 AAGACTTCGGCGCACCAGGATCCAGCATCTCAAGTGGGTGGTGCAGCCACGAGCT 195
DB 1 AAGACTTCGGCGCACCAGGATCCAGCATCTCAAGTGGGTGGTGCAGCCACGAGCT 60


```
QY 196 GCAGCAGGAGCAGCAGAGCAAGCAAGCCAGGGAACAGAGGAGGAGGCCGGGAGCAGCGACC 255
Db |||||||
61 GCAGCAGGAGCAGCAGAGCAAGCAAGCCAGGGAACAGAGGAGGAGGCCGGGAGCAGCGACC 120
QY 256 TGAGCAGCAGGCGAAGATGCTGAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGCGAGC 315
Db |||||||
121 TGAGCAGCAGGCGAAGATGCTGAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGCGAGC 180
QY 316 TTGTGACATCGCGTCAAGGTGTGCAGGCCCTACCTCTGGGCGAGCTGGAGCATGAGGACA 375
Db |||||||
181 TTGTGACATCGCGTCAAGGTGTGCAGGCCCTACCTCTGGGCGAGCTGGAGCATGAGGACA 240
QY 376 TCGACACGTGAGGAGGAGCTTCACTGAGGCCGAGTGGGAGGACCTTGACCC 435
Db |||||||
241 TCGACACGTGAGGAGGAGCTTCACTGAGGCCGAGTGGGAGGAGCTTGACCC 300
QY 436 AGCAGTACTACTCCCTCGTTCATGGCGATGCTTCACTCCAAATTCAGAAATTACTTTT 495
Db |||||||
301 AGCAGTACTACTCCCTCGTTCATGGCGATGCTTCACTCCAAATTCAGAAATTACTTTT 360
QY 496 CGCAGTGCCAGGCTCTGCTGAATAGAAATCAAGTCTGTGAACCTCAGACGGACATTTGAGC 555
Db |||||||
361 CGCAGTGCCAGGCTCTGCTGAATAGAAATCAAGTCTGTGAACCTCAGACGGACATTTGAGC 420
QY 556 GGCTCCGGAACATCTGGATTATAAGCCCGCGCCAAAGTCCCGGGGCCGAGACATATGTGT 615
Db |||||||
421 GGCTCCGGAACATCTGGATTATAAGCCCGCGCCAAAGTCCCGGGGCCGAGACATATGTGT 480
QY 616 GCATGACCGTGTGGAGGAGATCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCAGGG 675
Db |||||||
481 GCATGACCGTGTGGAGGAGATCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCAGGG 540
QY 676 ACAACAAAGTGGTGTTCACAGAGTATCATCGAGACGCGTCTCATCTGTGACACCAAGT 735
Db |||||||
541 ACAACAAAGTGGTGTTCACAGAGTATCATCGAGACGCGTCTCATCTGTGACACCAAGT 600
QY 736 TCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTCGACCATCTGCTTCTACA 795
Db |||||||
601 TCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTCGACCATCTGCTTCTACA 660
QY 796 AGGAGAGTTACTTGGGGTCTCAACTCAGCGCTTCTCCCTGCAGCAAGCTGGACAGGCCA 855
Db |||||||
661 AGGAGAGTTACTTGGGGTCTCAACTCAGCGCTTCTCCCTGCAGCAAGCTGGACAGGCCA 720
QY 856 TCACCTGTGCAACACGCGCTCCAGAGTACTGAAGAAATGATGTGGCGCGCAGCCGCC 915
Db |||||||
721 TCACCTGTGCAACACGCGCTCCAGAGTACTGAAGAAATGATGTGGCGCGCAGCCGCC 780
QY 916 TGCTGCCCGCACACAAATGTGGACACGAGTTCAGGAGTACCTGACGCGCCAGG 975
Db |||||||
781 TGCTGCCCGCACACAAATGTGGACACGAGTTCAGGAGTACCTGACGCGCCAGG 840
QY 976 GCGGTGGCGCGTGTGGGCGAGGTCTATCTACCGTCTGATGAAGAGGCGCATTCGCCACG 1035
Db |||||||
841 GCGGTGGCGCGTGTGGGCGAGCGTCTATCTACCGTCTGATGAAGAGGCGCATTCGCCACG 900
QY 1036 CCATGAAGGTGGCCAGGACACGCTGGAGCCTTCGCAAGAAAGCTTTGAGCTCTACGGGG 1095
Db |||||||
901 CCATGAAGGTGGCCAGGACACGCTGGAGCCTTCGCAAGAAAGCTTTGAGCTCTACGGGG 960
QY 1096 CTGACTTCGTCTCTGGAGGAGGACTTCAGGCCCTTGGCTGATCGAGATCAATTCAGGCCCA 1155
Db |||||||
961 CTGACTTCGTCTCTGGAGGAGGACTTCAGGCCCTTGGCTGATCGAGATCAATTCAGGCCCA 1020
QY 1156 CATGACACCGCTTCAGCCCGGTTCACGGCCCGAGTGTGTGACAGGTGAGGAGGACCA 1215
Db |||||||
1021 CATGACACCGCTTCAGCCCGGTTCAGCCCGAGTGTGTGACAGGTGAGGAGGACCA 1080
QY 1216 TCAAGGTGGCGGTGACCGCAGCTGTGACATCGGCAACTTCAGAGCTTCCTGTGAGGCGAGC 1275
Db |||||||
1081 TCAAGGTGGCGGTGACCGCAGCTGTGACATCGGCAACTTCAGAGCTTCCTGTGAGGCGAGC 1140
```

```
QY 1276 CGGTGGTTGAGCGCGGCCCAATTCAGCGGTCGACACTCTGCGTGGCGGCGTCAGT 1331
Db |||||||
1141 CGGTGGTTGAGCGCGGCCCAATTCAGCGGTCGACACTCTGCGTGGCGGCGTCAGT 1196

RESULT 2
AY415399 1257 bp DNA linear GSS 12-DEC-2003
LOCUS Pan troglodytes HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY415399
VERSION AY415399.1 GI:39771358
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 1257)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1257)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1257
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<!.>1257
/locus_tag="HCM5544"
gene
ORIGIN
Query Match 62.3%; Score 1145.6; DB 9; Length 1257;
Best Local Similarity 96.2%; Pred. No. 2.7e-241;
Matches 1151; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 136 AAGACTTCGGGCGCACCATGGCATCCAGTCATCCAGTGGGTGGTTCAGCCACCGAGCT 195
Db |||||||
1 AAGACTTCGGGCGCACCATGGCATCCAGTCATCCAGTGGGTGGTTCAGCCACCGAGCT 60
QY 196 GCAGCAGAGCAGCAGAGCAAGCCAGGGACACAGAGGAGAGGCCGGGAGCAGCGACC 255
Db |||||||
61 GCAGCAGAGCAGCAGAGCAAGCCAGGGACACAGAGGAGAGGCCGGGAGCAGCGACC 120
QY 256 TGAGCAGCAGGCAAGATGCTGAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGCGAGC 315
Db |||||||
121 TGAGCAGCAGGCAAGATGCTGAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGCGAGC 180
QY 316 TTGTGGAACATCGCGTGAAGGTGTGCCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACA 375
Db |||||||
181 TTGTGGAACATCGCGTGAAGGTGTGCCAGGCTTACCTGGGGCAGCTGGAGCATGAGGACA 240
QY 376 TCGACACGTGAGGAGGAGCTTCACTGAGGCCGAGTGGGAGGAGCTTGACCC 435
Db |||||||
241 TCGACACGTGAGGAGGAGCTTCACTGAGGCCGAGTGGGAGGAGCTTGACCC 300
QY 436 AGCAGTACTACTCCCTCGTTCATGGCGATGCTTCACTCCAAATTCAGAAATTACTTTT 495
Db |||||||
301 AGCAGTACTACTCCCTCGTTCATGGCGATGCTTCACTCCAAATTCAGAAATTACTTTT 360
QY 496 AGCAGTACTACTCCCTCGTTCATGGCGATGCTTCACTCCAAATTCAGAAATTACTTTT 360
Db |||||||
301 AGCAGTACTACTCCCTCGTTCATGGCGATGCTTCACTCCAAATTCAGAAATTACTTTT 360
```



```
QY 496 CGCAGTCCAGGCTCTGTAATAGAAATCAGTCTGTGAACCCCTCAGACGGACATTGACG 555
Db 361 CGNANNCCAGGCTCTCTGTAANNNTTCAANNNTNNNNCTNAGACGGACATTGACG 420
QY 556 GGCCTCCGGAACATCTGGATTATAAGCCCGCGCAAGTCCCGGGCGGAGACATAGTGT 615
Db 421 GGNCCGGAACATCTGGATTATAAGCCCGCGCAAGTCCCGGGCGGAGACATAGTGT 480
QY 616 GCATGGACCGTGTGGAGAGATCTCGAGCTGGAGCTGCAGACCAACCTCTTTCAGGG 675
Db 481 GCATGGACCGTGTGGAGAGATCTCGAGCTGGAGCTGCAGACCAACCTCTTTCAGGG 540
QY 676 ACAACAGTGGTGGTGCAGAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGT 735
Db 541 ACAACAGTGGTGGTGCAGAGTACATCGAGACGCGCTGCTCATCTGTGACACCAAGT 600
QY 736 TCGACATCAGACAGTGGTTCCTCGTACGGAATGGAACCCCTGACCATCTGTTCTTACA 795
Db 601 TCGACATCAGACAGTGGTTCCTCGTACGGAATGGAACCCCTGACCATCTGTTCTTACA 660
QY 796 AGGAGATTACTTGCCTTCTCAATCAGCGCTTCTCCCTGGACAGCTGGACGGCCA 855
Db 661 AGGAGATTACTTGCCTTCTCAATCAGCGCTTCTCCCTGGACAGCTGGACGGCCA 720
QY 856 TCCACCTGTGCAACACGCGTCCAGAGTACCTGAGAGATGATGTGGCGCGAGCCCC 915
Db 721 TCCACCTGTGCAACACGCGTCCAGAGTACCTGAGAGATGATGTGGCGCGAGCCCC 780
QY 916 TGCTGCCCGCACACAAATGTGGACAGCAGCTTCCAGGAGTACCTGCAGCGCCAGG 975
Db 781 TGCTGCCCGCACACAAATGTGGACAGCAGCTTCCAGGAGTACCTGCAGCGCCAGG 840
QY 976 GCGTGGCGCGTGTGGGAGCGTATCTACCGTCCATGAAAGGCGCATCGCCACG 1035
Db 841 GCGTGGCGCGTGTGGGAGCGTATCTACCGTCCATGAAAGGCGCATCGCCACG 900
QY 1036 CCATGAGTGGCGGACGACGCTGAGAGCTCGAGAGACGTTGAGCTTACGCGG 1095
Db 901 CCATGAGTGGCGGACGACGCTGAGAGCTCGAGAGACGTTGAGCTTACGCGG 960
QY 1096 CTGACTTCTGCTTGGAGGAGCTTACAGGCGCTGCTGATCGAGATCAATTCAGGCCCCA 1155
Db 961 CTGACTTCTGCTTGGAGGAGCTTACAGGCGCTGCTGATCGAGATCAATTCAGGCCCCA 1020
QY 1156 CCATGACCCGCTCAGCCCGTCAAGCCGACGCTGTGTGACAGGTCAGGAGGACACCA 1215
Db 1021 CCATGACCCGCTCAGCCCGTCAAGCCGACGCTGTGTGACAGGTCAGGAGGACACCA 1080
QY 1216 TCAAGTGGCGGTCGACGCTGACATCGCACTTCGAGCTCTGTGAGGCGAGC 1275
Db 1081 TCAAGTGGCGGTCGACGCTGACATCGCACTTCGAGCTCTGTGAGGCGGNN 1140
QY 1276 CGGTGGTTGAGCGCGCCCATTCAGCGGTCGCACTTCGCGTGGCGGGGCTCAGT 1331
Db 1141 NNGTGGTTGAGCGCGCCCATTCAGCGGTCGCACTTCGCGTGGCGGGGCTCAGT 1196

RESULT 3
LOCUS AK030151
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933401B17 product:hypothetical tubulin-tyrosine ligase containing protein, full insert sequence.
ACCESSION AK030151
VERSION AK030151.1 GI:26326134
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
```

```
JOURNAL MEDLINE
PUBMED 99279253
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE
PUBMED 20499374
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE
PUBMED 20530913
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE
PUBMED 11076861
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL MEDLINE
PUBMED 12420573
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL MEDLINE
PUBMED 12420573
REFERENCE 7
AUTHORS Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.jp, URL:http://genome.gsc.riken.jp/, tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
source Location/Qualifiers
1..2969
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4933401B17"
/db_xref="taxon:10090"
/clone="4933401B17"
/sex="male"
```

```
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.jp, URL:http://genome.gsc.riken.jp/, tel:81-45-503-9222, Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
Location/Qualifiers
1..2969
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4933401B17"
/db_xref="taxon:10090"
/clone="4933401B17"
/sex="male"
```



```

/tissue_type="testis"
/clone_lib="RIKEN full-length mouse cDNA library"
/dev_stage="adult"
490_2835
/notes="unnamed protein product; hypothetical
Tubulin-tyrosine ligase containing protein
(InterPro|IPR004344, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAC26811.1"
/db_xref="GI.26326135"
/translations="MSENGRKKVLSSTLSDGDKHEENLKKQGIPODLSSSPKLDRYKIA
RQUTERAKERKIFSYGHYPIVRLARLKGWKEKFFPKALQNLGSDSKSAETKE
NOETLALPRDFIDHVMSLVQNEIPLYVLTIRKDVVYHSITDQMLNHYGKLSAFST
KIGLCLNPSLSPVQVQANTPTFPRCYGKLTCESEKQFLDFRFTVAEILKMWVLHG
NYSKYKGSKKKEAKNSDPSPKDPENPDILKFLSQGVVDVACKVQYCAQLGLEHE
DIVSEASTALSEEWNLDOOYLLVHGNAISITDKSKVQCAQLLSKISVNPOT
EIDGRINATIKPAKASGRGDIVCMRVENILSLVDAQSQTTFKNNVQVQYIETPML
IYDTKPIRQWFLVTDWNPFLTWFYKESYFLAPSTFSLKLSAHLKNNISQRLK
NOKERSPLLCNNMTSTFVFLVKRGRGGTGGSIIVPSMKVATNMAVQAQDHEA
KNSFELYGADFLIGDRDFKFWLLIENSSTPMPSTPVAQLCAVQQEDTIKVVDRLK
DRNCDGNFLFMRQPAVEVLPFGNSDGLCEGIVISKAKQMPPIASVGLSESLDAP
PKVRSALAMETVIRPRTVTRQDKGRKAEKVLSTTMSKMPQDAEVRGRAPKIYAEV
NDYQVQKSGKGYTRVGSQKVPVGTILSAQHPALFAQTKMTQMTSSPPPTASGNH
RDSSPCPTIFVEFSLVTKNSQSRCTIQRACQQTGPI"

```

ORIGIN

Query Match	44.5%;	Score 818.8;	DB 3;	Length 2969;
Best Local Similarity	74.8%;	Pred. No. 2.5e-169;		
Matches 1075;	Conservative	0;	Mismatches 342;	Indels 21; Gaps 3;
Qy	130	TTTCAGAGACTTCCGGCGCACCATGCATCCAGCATCCTCAAGTGGTGGTTCAGCCACC	189	
Db	1082	TCRTGGATGACTTCCGGCGCAGGTGCTGCCAGCATCCTCAAGTGGTGGTTCCTTCAACC	1141	
Qy	190	AGAGCTGCAGCAGGAGCAGCAGAGCAAGCCACGGGACACGAGGAGGAGGCGCGGAGCA	249	
Db	1142	AGAACTACT-----GCAGCAAGTCAAGGGCAAGTAGAGAGGAGGAGCCCAAGNACA	1195	
Qy	250	GCAGCTGAGCAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAGGGGCTCTCCGG	309	
Db	1196	GTGACCCCGACCCCAAGAAAGATCTGAAAATCTGACCTCAAGCTACCAAGGCTCTCGG	1255	
Qy	310	GGCAGCTTGTGGACATCCGCTGCAAGTGTCACAGGCTACCTGGGCGAGCTGGAGCATG	369	
Db	1256	GACAGGTGTGAGACACAGCTTGCAGGTGTGCAGGCTACCTGGGACAGCTAGAGCATG	1315	
Qy	370	AGGACATTCGACAGTTCAGCAGATGCCGTGGAGGACCTCACTGAGGCGGAGTGGGAGGACC	429	
Db	1316	AAGACATAGAGTATCGAGGCCAGCACCCAGGCGCTTGTCTGAGGAGGAGTGAATGACC	1375	
Qy	430	TGACCCAGCAGTACTACTCCCTGTTTCATGGCGAGTCTTTTCATCTCOATTCAGAAATT	489	
Db	1376	TGACACAGCAGTACTATCTACTGGTTTCATGGCAATGCTTCCATCACTGATTCCGAAAGTT	1435	
Qy	490	ACTTTTCGAGTCCGAGGCTCTCGTAATAGAAATCAGTCTGTGTAACCTTCAGCAGGACA	549	
Db	1436	ACTTTGGCAATGCCAAGCTCTGCTGAGTAGAATCACTTCTGTGAAACCTTCAGACAGAGA	1495	
Qy	550	TTGACGGGCTCCGGAAACATCTGGAATTATAAGCCCGCGGCAAGTCCCGGGGCGGAGACA	609	
Db	1496	TCGATGGGATCCGGAAACATCTGATCATAAAGCCTGCAGCCAAGTCCCGGGGTCGAGATA	1555	
Qy	610	TAGTGTGCATGGACCGGTGTGGAGAGATCTCGAGCTGGCAGCTGCAGACCAACCTCTTTT	669	
Db	1556	TTGTGTGTATGGACCGGTGTGGAGAAATCTCTGAGTCTGTGGCGCGCAGACAGCACCA	1615	
Qy	670	CCAGGGACACAGTGGTGGTCCAGAAGTACATCGAGAGCCGCTGCTCATCTGTGACA	729	
Db	1616	CAAGGACAAATGGGTGGTCCAGAAGTACATCGAGACCCCAATGCTCATCTACGACA	1675	
Qy	730	CCAAGTTCGACATCAGACAGTGGTTCTCTGCTCAGGACTGGGAACCCCTCACCATCTGGT	789	

Db	1676	CCAAAGTTTGACATCAGACAGCTGGTTCTCGTCTCAGACTGGAATCCCCCTAACCAATCTGGT	1731
Qy	790	TTCTACAAGGAGAGTTTACTTTCGGGTTCTCAAATCTCAGCGCTTCTCCCTGGACAAGCTGGACA	849
Db	1736	TTCTACAAGAAAGCTACCTCGGCTTCTCCACAAGCGCTTCTCCCTGGACAAGCTGGACA	1795
Qy	850	GGCGCATTCACCTGTGTGCAACAAAGCGCGTTCAGAAAGTACCTGAAAGATGATGTGTGGCGCGCA	909
Db	1796	GGCGCATTCACCTGTGTGTAACAACTCCATCAGAGGGCGCTCAAAAATGATAAAGAGCGCA	1855
Qy	910	GGCCCCGTGTCGGCGCACAAATGATGGACACGACCCAGGTTTCCAGGAGTACTCGAGC	969
Db	1856	GTCCGCTGTACCTTTGTATAAATGTGGACCAAGCACTCGTTTCCAGGAGTACTCGCAGA	1915
Qy	970	GCCAGGCGCTGGCGCGTGTGGGGCAGCGTCACTACCCCGTCCATGAAGAAAGCCATCG	1029
Db	1916	AGAGGGGCCAGAGAGGNAAGTGGGGTAGCATCATCTACCCGCTATATGAAAAGAGCTCTCA	1975
Qy	1030	CCCAAGCCATGAAGGTGGCCCAAGACAAGTGGAGGCTTCGCAAGAAACAGCTTTTGAGCTCT	1089
Db	1976	CCAATGCATGAGGTAGCCAGGACCCACGTAGAAAGCCCGTAAGAACAGCTTTCGAGCTCT	2035
Qy	1090	ACGGGCTGACTTTCGTCCTTGGAGGGAGCTTTCAGGCGCTGGCTGATCGAGATCAATTTCCA	1149
Db	2036	ACGGAGCCGACTTCATCTCTGGGGCGAGACTTCAAGCCATGGCTTATCGAGATCAACTTCCA	2095
Qy	1150	GCCCCACATGCACCGCTCCACGCCGCTCACGCCACGCTGTGTGCACAGTGCAGGAGG	1209
Db	2096	GCCCCACATGCACCCCTCCACTCCCGTCAAGCTCAGCTCTGTGCCAGGTGCAGGAGG	2155
Qy	1210	ACACCATCAAGGTGGCGTGGACCGC-----AGCTGTGACATCGGCAACTTCG	1257
Db	2156	ACACCATCAAGGTGGTGGTGGATCGCAAACTGGACCGAAACTGTGACATCGGCAACTTTG	2215
Qy	1258	AGCTCTGTGGAGGACGCGGTGGTTGAGCGCGCCCAATTCAGCGGGTTCGACCTCTCGG	1317
Db	2216	AGCTTCTGTGGCGGACGCTTCGCGCTGGAGCTGCCACCGTTCAACGGGTCTGACCTCTGTG	2275
Qy	1318	TGGCGGCGTCAAGTGTGAGGAGAGCCAGGAGGAGGTGCTGCCCTCTGCAACTCAAGG	1377
Db	2276	TGGAAGGTATCAGGTGAGAAAGCCAAAGCAATTCGCCCAATTCGCAGGCTCGGCC	2335
Qy	1378	CCTCGGCTCGCTGTTTGGAGCGCAGCCGCTGAAGGCAAGCGGGGCCCTCGGCCATGCCAG	1437
Db	2336	TCTCAGAATCACTCTTAGATGC--TCCACCAAGTGGCGAGTGCCTCGGCGCTGTATGG	2392
Qy	1438	ACCGTCCCAAGGACCCCAATCAGAGCTCTCAGCGGGACTTGGGACTGAAGGAAGAGA	1497
Db	2393	AAACAGTGTATCAGGCTTCCAAAGAACACCGTGGCGAGGACTGGAACCGGGAAGAAAGCAA	2452
Qy	1498	AGGGGCTCCCGCTGGCTCTCTGCACCTTAAAGGGGGGAGCCGAGAGCGGTGGAGC	1555
Db	2453	AGGTACTTTCAACCACTTGTTCCTATGCCAGTAAATGATGCAGAGGTGCAGAGGTGCGAGC	2510
RESULT 4			
AK029745			
LOCUS	AK029745	2720 bp	mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930524K07 product:hypothetical tubulin-tyrosine ligase containing protein, full insert sequence.		
ACCESSION	AK029745		
VERSION	AK029745.1	GI:26325681	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)		
	99279253		

[illegible]

PUBLISHED REFERENCE AUTHORS	10349636 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE PUBLISHED	Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
PUBLISHED REFERENCE AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE PUBLISHED	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
PUBLISHED REFERENCE AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBLISHED	Nature 409, 685-690 (2001) 5
PUBLISHED REFERENCE AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL MEDLINE PUBLISHED	Nature 420, 563-573 (2002) 6 (bases 1 to 2720)
PUBLISHED REFERENCE AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Azakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saio, R., Saichoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL MEDLINE PUBLISHED	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/.
FEATURES source	Location/Qualifiers 1. .2720 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:4930524K07" /db_xref="taxon:i0090" /clones="4930524K07" /sex="male" /tissue type="testis" /clone_id="RIKEN full-length enriched mouse cDNA library"


```
QY 850 GCGCCATCCACCTGTGCAACAAACCGCGTCAGAAAGTACCTGAAGAATGATGTGGGCCGCA 909
Db 1799 GCGCCATCCACCTGTGTAAACAACTCCATCCAGAGCGCCTCAAAAATGATAAAGCGCA 1858
QY 910 GCCCCTGTGCGCGCACACAAATGTGGACACAGCAGCTTCCAGGAGTACCTGCAGC 969
Db 1859 GTCCGTGCTTACCTTGTATACATGTGGACCAAGCACTCGTTTCCAGGAGTACCTGCAGA 1918
QY 970 GCCAGGCGCGTGGCGCGTGTGGGGAGCGTCACTTACCCGTCCATGAAGAAGGCCATCG 1029
Db 1919 AGAGGGGCGGAGGAGAACTGGGGTAGCATCATCTACCGTCTATGAAGAAGCTGTCA 1978
QY 1030 CCACGCCATGAAGGTGGGCCAGACACAGTGAGGCTTCGAAGAAGACAGCTTTGAGCTCT 1089
Db 1979 CCAATGCCATGAGGTAGCCAGGACACGTAGAAGCCGTAGAAGCAGCTTCGAGCTCT 2038
QY 1090 ACGGGGCTGACTCTGCTTGGGAGGAGCTTACAGGCCCTGCGTGTGATCGAGATCAATTCCA 1149
Db 2039 ACGGAGCGGACTTCTCTGGGGCGAGACTTCAAGCCATGGCTTATCGAGATCAACTCCA 2098
QY 1150 GCCCCACCATGCAACCGCTCCACCGCGGTACAGGCCAGCTGTGTGCACAGGTGCAGGAGG 1209
Db 2099 GCCCCACCATGCAACCGCTCCACCTCCGTCACAGCTCAGCTGTGTGCCAGGTGCAGGAGG 2158
QY 1210 ACACATCAAGTGGCGG-----TGGACCGAGCTGTGACATCGGCAACTTCG 1257
Db 2159 ACACCATCAAGGTGGTGGTATCGCAAACTGACCGAAACTGTGACATCGGCAACTTTG 2218
QY 1258 AGCTCTGTGGAGCGAGCGGTGTGAGCGCCGCCATTCAGCGGCTCCGACCTCGG 1317
Db 2219 AGCTTCTGTGGCGGAGCGCTGCCGTGAGCTGCCACCGTTCAACGGGTCTGACCTCTGTG 2278
QY 1318 TGGCGGCGCTCAGTGTGAGGAGAGCCAGGAGGAGGTGCTGCCCTGTGCAACTCAAGG 1377
Db 2279 TGGAGGTATCAGCTGAAGAGAGCCAGAGCAATGCCGCCATGTCAGCGTCGGC 2338
QY 1378 CTTGGCGCTGCTGTGTGAGCGCAGCGCTGAAGGACAGGGGCCCTCGGCCATGCCAG 1437
Db 2339 TCTCAGAACTACTCTTTAGATGC---TCCACCCAAAGTGGCGAGTGGCGGCGCTGATGG 2395
QY 1438 ACCCTGCCAGGAGACCCCAATCACAGCTCTCCAGCGGAGCTTGGGACTGAGGAGAGAGA 1497
Db 2396 AATACGTATCAGGCTTCCAAAGAACCGCTGCGGAGGACTGGAACCGGAGAGAGCAA 2455
QY 1498 AGGGGCTCCCGCTGGCTGCTGGCACCCCTTAAGGGGGGAGCCGAGAGCGGTGGAGCGG 1557
Db 2456 AGGGCTTCGCTGTGATCTCCAAGCGCGTGCCTAGTTTGTGTGGCTTCGCGCGCGCG 2515
QY 1558 CACAGCCCAACCGCACCAAGCTGTCTGGGAGGTGAGCTCCCGGCTCGCCCTGTGTGC 1616
Db 2516 CCCTGGCCCTGCAAGGCTCTCCCTGCGCCTTGATGCGCGAGGGCGCTTCTGTGTGCTG 2574
```

```
RESULT 5
CR745100 770 bp mRNA linear EST 30-AUG-2004
LOCUS CR745100 Soares testis NHT Homo sapiens cDNA clone IMAGE97102270 ;
DEFINITION IMAGE:1645239 5', mRNA sequence.
ACCESSION CR745100
VERSION CR745100.1 GI:51667573
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 770)
Peters,M., Radelof,U. and Schneider,D.
JOURNAL I.M.A.G.E. cDNA Clone Collection
COMMENT Unpublished (2004)
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
```

Email: www.rzpd.de
RZPD; IMAGE97102270.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACAGGAACAGCTATGAC.

FEATURES
source

Location/Qualifiers
1..770
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE97102270 ; IMAGE:1645239"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTTACCAATCTGAAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaudo."

ORIGIN

	Query Match	41.8%;	Score 768;	DB 7;	Length 770;
	Best Local Similarity	99.7%;	Pred. No. 3.1e-158;		
	Matches 768;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	403	ACCTCACTGAGCGCGAGTGGAGGAGCCTGACCCAGCAGTACTACTCCCTCGTTCATGGCG	462		
Db	1	ACCTCACTGAGCGCGAGTGGAGGAGCCTGACCCAGCAGTACTACTCCCTCGTTCATGGCG	60		
QY	463	ATGCTTTTCATCTCCAATTCAAGAAATTTACTTTTCGAGTGCAGGCTCTGCTGAATAGAA	522		
Db	61	ATGCTTTTCATCTCCAATTCAAGAAATTTACTTTTCGAGTGCAGGCTCTGCTGAATAGAA	120		
QY	523	TCACCTCTGTGAACCTCAGACGGACATTCAGCGGCTCGGAAACATCTGGATTATAAAGC	582		
Db	121	TCACCTCTGTGAACCTCAGACGGACATTCAGCGGCTCGGAAACATCTGGATTATAAAGC	180		
QY	583	CCGCGGCCAAAGTCCCGGGGCCGAGACATAGTGTGCATGGACCGTGTGGAGGAGATCCTGG	642		
Db	181	CCGCGGCCAAAGTCCCGGGGCCGAGACATAGTGTGCATGGACCGTGTGGAGGAGATCCTGG	240		
QY	643	AGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGAACAAGTGGGTGTCAGAGATACA	702		
Db	241	AGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGAACAAGTGGGTGTCAGAGATACA	300		
QY	703	TCGAGACCGCGCTGCTCATCTGTGACCAAGTTCGACATCAGACAGTGGTTCCTCGTCA	762		
Db	301	TCGAGACCGCGCTGCTCATCTGTGACCAAGTTCGACATCAGACAGTGGTTCCTCGTCA	360		
QY	763	CGGACTGGAACCCCTCGTGCATCTGTTCTACAAGGAGAGTTTACTTTGGGTTTCTCAACTC	822		
Db	361	CGGACTGGAACCCCTCGTGCATCTGTTCTACAAGGAGAGTTTACTTTGGGTTTCTCAACTC	420		
QY	823	AGCGCTTCTTCCCTGGACAAAGCTGGAAGCCCATCCACTGTGCAACAACCGCGTCCAGA	882		
Db	421	AGCGCTTCTTCCCTGGACAAAGCTGGAAGCCCATCCACTGTGCAACAACCGCGTCCAGA	480		
QY	883	AGTACTCTGAAGAATCATGTGGGCCCGACCCCTCTGCCCGCACACACATGTGGACCA	942		
Db	481	AGTACTCTGAAGAATCATGTGGGCCCGACCCCTCTGCCCGCACACACATGTGGACCA	540		

RESULT #	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Bu937372		BU937372	775 bp	mRNA	linear	EST 18-OCT-2002					
AGENCOURT	10519180	NIH MGC 169	Mus musculus	cdna clone							
IMAGE:	6705256	5', mRNA sequence.									
BU937372		BU937372									
BU937372.1	GI:24126191										
EST.											
Mus musculus											
Mus musculus											
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.											
1 (bases 1 to 775)											
NIH-MGC	http://mgs.nci.nih.gov/.										
National Institutes of Health, Mammalian Gene Collection (MGC)											
Unpublished (1999)											
Contact: Robert Strausberg, Ph.D.											
Email: rs@bbs-r@mail.nih.gov											
Tissue Procurement: Dr. Jonathan Kuo, NIMH											
cdna Library Preparation: Michael Brownstein Laboratory											
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)											
DNA Sequencing by: Agencourt Bioscience Corporation											
Clone distribution: MGC clone distribution information can be											
found through the I.M.A.G.E. Consortium/LLNL at:											
http://image.llnl.gov											
Plate: LCM3002	row: 9	column: 16									
High quality sequence	stop: 569.										
Location/Qualifiers											
1..775											
/organism="Mus musculus"											
/mol_type="mRNA"											
/db_xref="taxon:10090"											
/clone="IMAGE:6705256"											
/lab_host="DH10B (TI-phage-resistant)"											
/clone_lib="NIH MGC 169"											
/note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccctcgcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCATGAGTATCAACGACGAGTGGCCATTCGCGGG-3' and 5'-ATTCTAGAGCCGAGCGGCGGACAT-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."											
ORIGIN											
Query Match	24.1%	Score	443.6;	DB 5;	Length	775;					
Best Local Similarity	74.5%;	Pred. No.	7.7e-87;								
Matches	576;	Conservative	0;	Mismatches	185;	Indels	12;	Gaps	1;		
QY	738	GACATCAGACAGTGGTTCTCGTCACGGATCGAACCCCTGACCATCTCGTTCTACAAG	797								
Db	1	GGCATCAGACAGTGGTTCTCGTCACAGACTGGAATCCCTTAACCATCTCGTTCTACAAA	60								
QY	798	GAGAGTTTACTTTCGGGTTCTCAACTCAGCGCTTCTCCCTCGACAAGCTGGACAGCGCCATC	857								
Db	61	GAAGGTACTCTCGTTCTCCACACAGCGGTTCTCCCTCGACAAGCTGGACAGCGCCATC	120								
QY	858	CACCTGTGCAACACGCCGCTCCAGAGTACTCGAGGAGGCGCATTCGCCACGCC	917								
Db	121	CACCTGTGTAAACAATCCATCCAGAGCGCGCTCAAAAATGATAAAGAGCGAGTCCGCTG	180								
QY	918	CTGCGGCGCCATGTGGGCGACGCTCATCTACCCGCTCCATGAAGAGGCGCATTCGCCACGCC	977								
Db	181	CTACCTTGTATTAACATGTGGACGAGCACTCTGTTTCCAG									


```

/notes="hypothetical protein"
1..4184
/gene="DKFZp686D076"
2529..3587
/gene="DKFZp686D076"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH10554.1"
/db_xref="GI:50949619"
/translation="MDIDKLEAPLYLTPGWSLFLQRYQVYVHGEALRLHLDLTVQVR
CEDLQQLQAVPQIDMEGRNTWI VKPGAKSRGIMCMHLEMLKLVGNPNVYMK
DGKVVQVQLLI FETGFLRQWFLVDWNPFLVWFYRDSYIRFSTQFSLKMLD
NSVHLNNSIOKLENSCHRPILLPPDNMSSORFOAHLQEMGAPNAWSTIIYVPGMKD
AVIHALTSQTVQCRKASRELVCADVFQEDFQPLIEINASPTWAPSTAVTABLCA
GVQADTLRVVIDRMLDRNCTGAFELLYKQPVITSPASTPGPSCLLPIMYSDTRRSSD
DSTASWWALRPCRQARP"

Query Match      22.9%; Score 421.4; DB 3; Length 4184;
Best Local Similarity 63.8%; Pred. No. 7.5e-82;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;

Qy 232 GGGAGGAGCGCGGAGCAGCGACCTGAGCAGCAGGCAAGATGCTGAAAAATGCTGAGGCAA 291
Db 2391 GGGATCGTAGAGGCCCTCCCTATGTGACAGGACAAAGCAGCCCAAGAAACAGGAGAA 2450
Qy 292 AGCTCAG- GGGCCTCCCGGGGACGCTGTGTGACATCGCGGTGCAAGGTGTGCGAGGCCCTAC 350
Db 2451 AACCCAGTGTGTGTCCTCCCGCAGAGTTTGTGTGATGAAGCTCTGTGTGCGTGCAGGAGTAC 2510
Qy 351 CTGGGGCAGCTGGAGCATGAGGACATGACACGCTGACAGATCGCGTGGAGGACCTCACT 410
Db 2511 CTTAGCAACTTGTGGCCCATGATGACAGGACCTGAGGCGCCGCTGTACTCTCACC 2570
Qy 411 GAGCGCAGTGGGAGGACCTGACCCAGCAGTACTCTCCCTGCTTCATGCGCATGCTTTC 470
Db 2571 CCCGAGGCTGGTCCCTCTTCTCCAGCGTACTACTACCAAGTGTCCAGAGGGGAGAA 2630
Qy 471 ATCTCCAAATCAAGAAATTAATTTTTCGAGTGCCAGGCTCTGTGTAATAGAAATCACTCT 530
Db 2631 CTCAGGCACCTCGACACTCAGGTCAGCGCTGTGAGGACATCTCTGACAGCAGCTGCAGGCC 2690
Qy 531 GTGAACCTCAGACGACATGAGGCTCGGAACATCTGGATTATTAAGCCCGCGGCC 590
Db 2691 GTGTATCCCAAGATAGACATGAAAGGGGATCGCAACATCTGGATCGTGAAGCAGGAGCC 2750
Qy 591 AAGTCCCGGGCCGAGACATAGTGTGATGACCGCTGTGGAGGAGATCTCTGGAGCTGGCA 650
Db 2751 AAGTCCCGGGCAGGACATCATGTGATGGACCACTCTGGAGGAGATGCTGAAGCTGGT 2810
Qy 651 GCTGCAGACCACTCTTTTCCAGGGAACAAGTGGGTGTGTCAGAAAGTACATCGAGACG 710
Db 2811 AACGGCAACCCGTGGTGTATGAAGGACGGCAAGTGGGTGTGTCAGAAAGTATATTGAGCGG 2870
Qy 711 CCGTGTCTCATCTGTGACACAGTGTGACATCAGACAGTGTTCCTCGTCAGGACTGG 770
Db 2871 CCCCCTCTCATCTTTTGGCAACAAGTTTGACTCAGACAGTGTTCCTGGTAACTGACTGG 2930
Qy 771 AACCCCTGTGACCATCTGTTCTCAAGAGAGATTACTTGGGGTGTCTCAACTCAGCGCTTC 830
Db 2931 AACCCACTTACCGTGTGGTCTACCGGACAGCTATATCGCTTTTCCAGCGAGCCCTTC 2990
Qy 831 TCCCTGACCAAGCTGGACAGCGCCATCACTGTGCAACACCGCGTCCAGAAAGTACTCTG 890
Db 2991 TCCCTGAAGAACCTGGACAACTCAGTGCACCTGTGCAACAACTCCATCCAGAAAGCACTG 3050
Qy 891 AAGATCATGTGGCCCGCAGCCCTCTGTGTCGCCGACACACATGTGTGACCGACGACGAGG 950
Db 3051 GAGAACTCATGTGCCATCGGCCATCCACTGCTTTCGGCAGACAACTATGTGGTCTAGCAGAGG 3110
Qy 951 TTCCAGGAGTACCTTGACGAGCGCCGCGGTGGCGCGTGTGGGCGCAGCGTCACTACCCG 1010
Db 3111 TTCCAGGCCCACTGACAGGAGTGGGTGGCGCCCAATGCTTGGTCCACCATCATCTGTCCT 3170
```

ORIGIN

```

Qy 1011 TCCATGAGAGGCCATGCCCCAGCCATGAAGTGGCCAGCACACGCTGGAGCCTCGC 1070
Db 3171 GGCATGAAGGATGCTGTGATCCAGCACAATTCAAGACCTCCAGGACACCGTGCAGTGTGCG 3230
Qy 1071 AAGAAACAGCTTTGAGCTCTACGGGGGTGACTTCGTCTCTTGGGAGGAGACTTTCAGGCCCTGG 1130
Db 3231 AAGCCAGCTTTGAGCTCTATGGCGGTGACTTCGTGTTCGGGGAGGACTTCCAGGCCCTGG 3290
Qy 1131 CTGATCGAGATCAATTCACGCCCCACCATGACCCGCTCCAGCGGTTCAGGCCCCAGCTG 1190
Db 3291 CTGATTCAGATCAACGCCAGCCCAAGATGGCAACCTCCACAGCAGTCACTGTCCCGGCTC 3350
Qy 1191 TGTGCACAGCTGCAGGAGGACACATCAAGTGGCC-----GTGGACCGCAGC 1238
Db 3351 TGTGCTGGCGTGAAGCTGACACCTCGCGGTGTTCATTACCGGATGCTGGACCGCAAC 3410
Qy 1239 TGTGACATCGGCAACTTCGAGCTCCTGTGGAGGAGCGCGGT 1279
Db 3411 TGTGACACAGGAGCCTTTGAGCTCATCTATAAGCAGCCCT 3451

RESULT 9
CK597795 777 bp mRNA linear EST 22-JAN-2004
AGENCOURT 17837841 NIH_MGC_238 Rattus norvegicus cDNA clone
IMAGE:7135519 5', mRNA sequence.
CK597795
CK597795.1 GI:41110910
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 777)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg 31 Rm10A07 Bethesda, MD 20892
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Howard Jacobs
CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM15042 row: 0 column: 05
High quality sequence stop: 693.
FEATURES
Location/Qualifiers
1..777
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7135519"
/tissue_type="testis, pooled"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_238"
/notes="Organ: testis; Vector: pExpress-1; Site 1: EcorV;
Site 2: NotI; RNA obtained from testis tissue of 8 wk old
animal. Tissues were snap-frozen and kept at -80C before
RNA extraction and purification (Tri-reagent method). CDNA
was primed using oligo-dT primer:
5'-pGACTAGTCTAGATCGAGCGCCGCCCT(T)25-3' and cloned into
the EcorV/NotI sites of pExpress-1. Size-selection >1.4kb
resulted in an average insert size of 1.9 kb. This primary
library is normalized (non-normalized primary library is
NIH_MGC_237) and was constructed by Express Genomics
(Frederick, MD)"
```


ORIGIN	10349636	2	PUBMED	REFERENCE	AUTHORS
Query Match	22.4%;	Score 412.4;	DB 7;	Length 777;	
Best Local Similarity	77.0%;	Pred. No. 5.6e-80;			
Matches 530;	Conservative	0;	Mismatches 151;	Indels 7;	Gaps 2;
QY	130	TTTCAGAAGACTTCCGGCGCACCATCGCATCTCTCAAGTGGTGGTTCAGCCACC	189		
DB	19	TCTTGGATGACTTCCGGCGTACAGTGGCTGCCAGCATCTCTCAAGTGGTGGTTCATCACC	78		
QY	190	AGAGCTCGCAGGAGCAGCAGAGCAAGCCAGGACAGAGGAGGAGCGCGGAGCA	249		
DB	79	AGAACTACTGCAA-----CAGAGCCAAAGGCAAGAGTAGAAGAGGAGGCGCAAGATG	132		
QY	250	GCACCTCGAGCAGCAGCGCAAGATGCTGAAATGCTCAGGCAAGCTCAGGGGCTCCCG	309		
DB	133	GTGACCCCGCCCAAGAAGATCTTGATATCTGACCTCAAGTTACCAAGCCTCTCAG	192		
QY	310	GGCAGCTTGTGGACATCGCGTGAAGGTGTCAGGCGCTACCTGGGGCAGCTGGAGCATG	369		
DB	193	GACAGGTTGTAGACACAGCATGCAAGGTGTGCCAGGCTACCTGGGACAGCTGGAGCATG	252		
QY	370	AGGACATCGACAGCTCAGCAGATGCCGTGGAGACCTCACTGAGCGGAGTGGAGGCC	429		
DB	253	AGGACATCGACCTGTGAGACGCGAGCGGAGGCGCTTGTCCGAGGAGGAATGGAACGCC	312		
QY	430	TGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTCATCTCCAATTCAAGAAAT	489		
DB	313	TGACACAACAATACTACTTCTGTTTCATGGCNAATGCTTCCATCAGGATTCAGNAGTT	372		
QY	490	ACTTTTCGAGTGCCAGGCTGTGCTGAATAGAAATCAGCTGTGTAACCTCTCAGACGACA	549		
DB	373	ACTTTGCGCAGTGCCAAAGCTGTCTGAGTAAGATCAGCTCTGTGAACCCCGACCGAGA	432		
QY	550	TTGACGGGCTCCGGACATCTGATTAATAAGCCCGCGGCGGAGTCCCGGCGCGAGCA	609		
DB	433	TTGACGGGATACGGAACATCTGGATCATTAAGCCTTCAGCCCAAGTCCCGGGCGGAGATA	492		
QY	610	TAGTGTGCATGGACCGTGTGGAGGAGATCTGTGAGCTGGCAGCTGCAGACCAACCTCTTT	669		
DB	493	TTGTGTGCATGGACCGTGTGGAGAAATCTCTGGATCTGTGTGGCCACAGACGACGACA	552		
QY	670	CCAGGACAACAAGTGGGTGCTCCAGAAGTACATCAGACGCGCGTGTCTATCTGTGACA	729		
DB	553	CGAAGGACAACAATGGGTGCTCCAGAAGTACATCAGACCCCGGATGCTCATCTATGACA	612		
QY	730	CCAAGTTCGACATCAGACAGTGTCTCTCAGGACTGGAACCCCTGACCATCTGCT	789		
DB	613	CAAAGTTTGACATCAGACAGTGTCTCTTGTTCAGACATGGAAATCCCTT-AAACATCTGCT	671		
QY	790	TCTACAAGGAGAGTTACTTTCGCGTTCTC	817		
DB	672	TCTACAAGGAGTACTCTCGGTCTCAC	699		
RESULT 10					
AK080321					
LOCUS	AK080321	2520 bp	mRNA	linear	HTC 03-APR-2004
DEFINITION	Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630053H17 product:HO1TL PROTEIN homolog [hom sapiens], full insert sequence.				
ACCESSION	AK080321				
VERSION	AK080321.1	GI:26348476			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
REFERENCE	1				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Hirozane, T., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				
PUBMED	11076861				
REFERENCE	4				
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409, 685-690 (2001)				
REFERENCE	5				
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				
REFERENCE	6				
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)				
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ Location/Qualifiers 1. 2520 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:A630053H17" /db_xref="taxon:10090"				
FEATURES	source				


```
/clone="A630053H17"
/tissue="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
84.72198
/note="unnamed protein product; HOTT1 PROTEIN homolog
[ Homo sapiens ] (spt8) [ Q9U199, evidence: FASTV, 85.8% ID,
93.4% length, match=821 ]
putative"
/codon_start=1
/protein_id="BAC37878.1"
/db_xref="GI:26348477"
/translation="MDTSSPFYAAGQGEASPPSLPVSISPPSLQSRMVRNETPVLWT
TRDVLDRSLSDQMLNHYARAGSTTKVGLCLNLRNLPFDEADADSPFPCYRLG
AEDKKAIFEDFHLTAARNVLKLVKLEKSQISIQAREEAPEDTPKPKKLVLT
VSSDFVDEALSCQHLNLSIAHKDIIDKPNPILYLPDDWSQFLQRYQIVHGEALR
YLSDFQVQREDIIOQLNVPOLDMDGRDNIWIKPKAKSRGRGIMCMRLDMLKLV
DCNPMKMDGKWTVKIERPLLI FGTKFLRQWFLVTDWNPVPLVYFIRDSYIRFSTQ
PFLKLNLDNVHLNCSIORHLEASCHRRPMLPDNNWSORFOALHVEDAPKASQ
VIVPGKAAVIHALQTSQNVQCKASFELYGADVFGEDEFQPLMLEINASPSTAPST
AVTARLCAGVQATLVRIDRLDRSCDTGAFELIYKQPAVEVPQVYGLNLLVGGSTI
KQPVVGHRTTVRSLSLPHLLTQGGSKESQDSPTHRSSASRNKRAESLEHTEKEPEP
AAVASGKGKAPFPFSLHSAWLPSPRVRHPQGRVLRQLDQLVGSKALSTTKGA
LMTLPKVLMSPPHPDLKAPMLKPKGVGFELCCTTWRVVLSSGIGSEGRQRAA
PRPSAPGKLSSTPCSTET"
```

ORIGIN

Query Match 21.7%; Score 399; DB 3; Length 2520;
Best Local Similarity 61.8%; Pred. No. 5.9e-77;
Matches 650; Conservative 0; Mismatches 395; Indels 12; Gaps 1;

```
QY 291 AAGCTAGGGGCTCCCGGGCAGCTTGTGCAGATCGCGTCAAGTGTGCCAGGCTAC 350
Db |||||
QY 552 AAGCTAGTACCGGTCTCTCAGACTTTGTGCGATGAGGCTCTGATGCATCCAGGAGCAC 611
Db |||||
QY 351 CTGGGGCAGCTGAGCATGAGGACATCGACACGTCAGCAGATCGCGTGGAGGACCTCACT 410
Db |||||
QY 612 CTTAGCAGCATAGCCCAAGACATTTGACAGAGACCCCAACTCCCGCTCTACTCAGC 671
Db |||||
QY 411 GAGCGCAGTGGGAGGACCTGACCCAGCAGTACTCTCCCTCGTTCAATGCGGATGCTTTC 470
Db |||||
QY 672 CCGCATGATTGGTCCAGTTCCTTCAGCGCTACTTACCACAAATAGTTTCATGAAGGGCAGAG 731
Db |||||
QY 471 ATCTCCAAATCAAGAAATTAATTTTGCAGTGCAGGCTCTGCTGATAGATCAATCAGTCT 530
Db |||||
QY 732 CTGATACCTGAGGTCCAAGTCCAGCGCTGTGAAGACATCTCTACAGCAGTTCAGAAC 791
Db |||||
QY 531 GTCAACCTTCAGACGGGCTTTCAGCGGCTCCGGAACATCTGGATTATAAGCCCGCGGCC 590
Db |||||
QY 792 GTGTACCCAGTTAGACATGAGGGGGATCGGAACATCTGGATCTGAAGCCCGGAGCC 851
Db |||||
QY 591 AAGTCCCGGGCCGAGACATAGTGTGCATGAGACCGTGTGGAGGAGATCTCTGGAGCTGGCA 650
Db |||||
QY 852 AAGTCCCGGCGCGAGGATTATGTGTCATGAACCGCTGATGAGATGCTGAAGCTGGTG 911
Db |||||
QY 651 GCTGCAGACCACTCTTTCAGGAGCAACAAGTGGTGTGTCAGAGTACATCGAGAG 710
Db |||||
QY 912 GACTGCAACCCCATGCTCATGAAGGATGGCAAGTGGATCGTGCAGAGTACATTTGACGGG 971
Db |||||
QY 711 CCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGG 770
Db |||||
QY 972 CCCCTGCTTAATCTTGGCACAAGTTTGACTGAGACAGTGGTTCCTGTTGACTGACTGG 1031
Db |||||
QY 771 AACCCCTGACCAATCTGGTTTCTAAGGAGAGTACTTGGCGTTTCTCACTCAGCGCTTC 830
Db |||||
QY 1032 AACCCACTACCGGTGTGTTTCTACCGGAGACAGCTACATTCGCTTCTCCACACAGCCTTC 1091
Db |||||
QY 831 TCCCTGCAAGCTGGACAGGCACTCCACTGTGCAACAACCGCGTCCAGAGTACCTG 890
Db |||||
QY 1092 TCCCTGAAGAACCTGGACAACTCTGTGCACTCTGTGAACAACTCCATCCAGAGACACTTG 1151
Db |||||
QY 891 AAGNATGATGGGCGCAGCCCCCTGCTGCCGCGACACACATGTGGACCAACCAACAGG 950
Db |||||
```

```
Db 1152 GAGGCTTCTGTCACTCCGGCACCCGATGCTGCCCCAGATAAATGTGGTCCAGCCAGG 1211
QY 951 TTTCCAGAGTACTCTGAGACCCCGAGGGCCGTGGCGCGTGTGGGCGACGTCATCTACCCG 1010
Db |||||
QY 1212 TTTCCAGCCCACTTGCAGGAGGTAGATGCCCAAGGCTTGTCCAGCGTCTGCTGCGCA 1271
Db |||||
QY 1011 TCCATGAAGAAGCCATCGCCACGCGCATGAAGTGGCCAGGACCAACGTTGGAGCCTCGC 1070
Db |||||
QY 1272 GGCATGAAGGCTGCTGTGATCCATGCCCTGCAGACCTCCCAAGATAACGTGCGAGTCCGG 1331
QY 1071 AAGAACAGCTTTTCTGAGCTCTACGGGGTGTGCTTCTCTGGGAGGAGTTCAGGCCCCCTGG 1130
Db |||||
QY 1332 AAGGCCAGCTTTTCTGAGCTCTATGGGGCAGACTTTGTGTTGGGGAAGACTTTCAGCCCCCTGG 1391
QY 1131 CTGATCAGATCAATTCAGCCCCCAGCATGCACCCGTCAGCGCGGTGTCAGGCGCCAGCTG 1190
Db |||||
QY 1392 TTGATGAATCAATGCCAGCCCCCAGCATGCACCTTTCACGCGCTGTCTACTGCGCCGCTC 1451
QY 1191 TGTGCAGAGTGCAGAGGAGACAC-----CATCAAGTGGCGGTGGACCGCAGC 1238
Db |||||
QY 1452 TGTGCGGTGTGCAGGAGATACCTTGTGCTGTGCTGATCGACCGCGCTTGGACCGCAGC 1511
QY 1239 TGTGACATCGGCAACTTCGAGCTCTTGTGAGGAGCAGCGGTGTTGAGCGCCCCCAATTC 1298
Db |||||
QY 1512 TGTGATACGGGAGCCCTTGTGAGCTCATATAAGCAGCCTGCTGTGGAGGTGCTCTCAGTAC 1571
QY 1299 AGCGGTTCGACCTCTGCTGCGGGCGGTGCTGAGGAGAGCC 1343
Db |||||
QY 1572 GTGGGTATCGGCTCTCTAGTGGAGGCTCTTACCATCAAGAAGCCC 1616
```

RESULT 11

BM808637

LOCUS

DEFINITION

5', mRNA sequence.

1023 bp mRNA linear EST 05-MAR-2002

AGENCOURT_6617786 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734232

BM808637

BM808637.1 GI:19125460

EST.

Homo sapiens (human)

Homo sapiens

Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1023)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Invitrogen

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM12739 row: d column: 09

High quality sequence start: 6

High quality sequence stop: 664.

Location/Qualifiers

1..1023

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5734232"

/tissue_type="hippocampus"

/lab_host="DH10B"

/clone_lib="NIH_MGC_124"

/notes="Organ: Brain; Vector: pCMV-SPORT6; Site 1: EcorV (destroyed); Site 2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcorV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is

FEATURES

source

normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

Cracking code 012.

Query Match	20.6%	Score 378.4	DB 4	Length 1023
Best Local Similarity	69.5%	Pred. No. 1.8e-72		
Matches	514	Conservative	0	Mismatches 226; Indels 0; Gaps 0

Qy	497	GCAGTGCAGGCTCTGTGTAATAGAAATCAAGTCTGTGTGAACCTTCAGACGGACATTTGACGG	556
Db	49	CGCTGTGAGGACATCTCTGCAGCAGCTCCAGGCGGTGTATCCCCAGATAGACATGGAAGG	108
Qy	557	GCTCCGGAACATCTGGATTAAAGCCCGCGGCCAAGTCCGGGGCGAGACATAGTGTG	616
Db	109	GGATCGCAACATCTGGATCGTGAAGCAGAGACCAAGTCCCGGAGCGAGGCATCATGTG	168
Qy	617	CATGGAACCGTGTGGAGGAGATCTCTGGAGCTGGCAGCTGCAGACCAACCTCTTTCCAGGGA	676
Db	169	CATGGACCACTGGAGGAGATGCTGAAGCTGGTGAACGGCAACCCCGTGGTGATGAAGGA	228
Qy	677	CAACAAGTGGGTGTCCAGAAGTACATCGAGACGCCGCTCTCATCTGTGACACCAAGTT	736
Db	229	CGGCAAGTGGGTGTGAGAAATATTTGAGCGGCCCTCTCATCTTTGGCACCAGTT	288
Qy	737	CGACATCAGACAGTGGTTCCTCGTCAGGACTGGAAACCCCTGACCATCTGGTTCACAA	796
Db	289	TGACCTCAGACAGTGGTTCCTGGTAACTGACTGGAAACCACTTACCGTGGTTCACCG	348
Qy	797	GGAGAGTTACTTGGCGTCTTCAACTCAGCGCTTCTCCCTGGACCAAGCTGCAGACGCCAT	856
Db	349	CGACAGCTATATCCGCTTTTCCAGCGAGCCCTTCTCCCTGAAGAACCTGGACAACTCAGT	408
Qy	857	CCACCTGTGCACAAACGCCCGTCCAGAAGTACCTGAAGAATGATGTGGGCGCAGCCCCCT	916
Db	409	GCACCTGTGCAACAACTCCATCCAGAAAGCACCTGGAGAACTCATGCGCATCCACT	468
Qy	917	GCTGCCCGCACAAACATGTGGACCAACACAGTTCACGAGGTTCACGAGGTACCTGAGCGCCAGGG	976
Db	469	GCTTCCGCCAGAACAAATGTGTTCTAGCCAGAGTTCACGAGCCCACTGAGAGATGGG	528
Qy	977	CCGTGGCGCGTGTGGGCGAGCGCTCATCTACCCGCTCCATGAAGAAGGCCATCGCCACGC	1036
Db	529	TGCCCAAAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGGATGCTGTGATCCAGCG	588
Qy	1037	CATGAAGTGGCCAGACACACAGCTGGAGCCTCGCAAGAACAGCTTTGAGCTCTACGGGGC	1096
Db	589	ACTTCAGACCTCCACAGACACCGGTGCAGTGTCCGAAGGCCAGCTTTGAGCTCTATGGCGC	648
Qy	1097	TGACTTCGTCCTTGGGAGGACTTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCCAC	1156
Db	649	TGACTTCGTTTGGGAGGAGCTTTCAGGCCCTGGCTGATCGAGATCAATTCAGGCCCCAC	708
Qy	1157	CATGCACCCGTCACGCCGCTCAGGCCCAAGCTGTGTGCACAGGTGCAGGAGACACCAT	1216
Db	709	GATGGCACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGCAAGCTGACACCT	768
Qy	1217	CAAGTGGCGGTGGACCGCA	1236
Db	769	GCAGGTGGTCAATTGACCGGA	788

RESULT 12

BM808516

LOCUS

DEFINITION

AGENCOURT_6617783 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734160

5', mRNA sequence.

ACCESSION

BM808516

BM808516.1 GI:19125339

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Db 520 TGCCCCAAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGATGCTGTGATCCACGC 579

Qy 1037 CATGAAGGTGGCCCGCAGGACACGCTGGAGCCTCGCAAGAA CAGCTTTTTCAGCTCTACGGGGC 1096

Db 580 ACTTCAGACCTCCAGGACACCGGTGCAGTGTGGAAGGCCAGCTTTTCAGCTCTATGGCGC 639

Qy 1097 TGACTTCGTCTTTGGGAGGACATTCAGGCCCTGGCTGATCGAGATCAATTCACGCCCCAC 1156

Db 640 TGACTTCGTGTTTGGGGAGGACATTCAGGCCCTGGCTGATTCAGATCAACGCCCGCCCCAC 699

Qy 1157 CATGACCCCGTCACGCCCGGTTCAGGCCCGCAGCTGTGTGCACAGGTGCAGGAGGACCAT 1216

Db 700 GATGGCACCCCTCACAGCAGTCTCATGGCAGCTGTGTGCTGGGTCGAGCTGACACCT 759

RESULT	13
BM717061	
LOCUS	BM717061
DEFINITION	693 bp mRNA linear EST 28-FEB-2002 UI-E-EJ0-ahk-e-05-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone UI-E-EJ0-ahk-e-05-0-UI.5' mRNA sequence.

KEYWORDS	EST.	ORGANISM
SOURCE	Homo sapiens (human)	
	Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

FEATURES
source
seq primer: M13 reverse.
Location/Qualifiers
1. .693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-B-EJO-ahk-e-05-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

```

AGATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

visual system, supported by national eye institute (NEI).

ORIGIN	Query Match Best Local Similarity Matches 473; Conservative	17.5%; Score 321; DB 4; Length 693; 69.1%; Pred. No. 7.2e-60; 0; Mismatches 200; Indels 12; Gaps 2;
QY	603	CGAGACATAGTGCATGGACCGGTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCC 662
DB	1	CGAGGCATCATGTGCATGACACACCTGGAGGAGATGCTGAAGTGGT-GAAACGACCCC 59
QY	663	CTCTTTTCCAGGACAAACAGTGGGTGTCAGAGATCATCGAGACGCCGTGTCTCATC 722
DB	60	GTGGTGATGAAGGACGGCAAGTGGGTGGTGCAGAGTATATTGAGCGGCCCTCTCTCATC 119
QY	723	TGTGACACCAAGTTTGACATCAGACAGTGGTTCCTCGTCAACGGACTGGAAACCCCTGACC 782
DB	120	TTTGGCACCAGTTTGACCTCAGACAGTGGTTCCTGGTAACCTGACTGGAAACCCACTTACC 179
QY	783	ATCTGGTCTTCAAGGAGAGTTACTTTCGGGTTCTCAACTCAGCGTTCCTCCCTGGACAAG 842
DB	180	GTGTGGTCTTACCGGACAGCTATATCCGCTTTTCCACGACGCCCTTCTCCCTGGAAGAAC 239
QY	843	CTGGACAGCGCCATCCACCTGTGCAACACGCCGCTCCAGAACTACTGAAGAAATGATGTG 902
DB	240	CTGGACAACTCAGTGCACCTGTGCAACAACTCAATCCAGAAACCTTGAGAACTCATGC 299
QY	903	GGCGCAGCGCCCTGTGTCGCCGACACAAACATGTGGACACAGCACAGGTTCCAGGAGTAC 962
DB	300	CATCGCATCCACTGCTTCGGCCAGACAAACATGTGCTTAGCCAGAGGTTCCAGGCCAC 359
QY	963	CTGACGCGCCAGCGCGTGGCGCGTGTGGGGCAGCGTCACTCAACCGTTCATGAAGAAG 1022
DB	360	CTGACGAGAGTGGGTGCCCAAATGTTGGTCCACCATCATCGTGCTGGCATGAAGGAT 419
QY	1023	GCCATCGCCACGCATGAGTGGCCAGGACCAACGTTGGAGCCCTCGCAAGAACAGCTTT 1082
DB	420	GCTGTGATCCAGCACTTCAGACCTCCCAAGCACACCGTGCAGTGTGGAAAGCCAGCTTT 479
QY	1083	GAGCTCTACGGGGCTGACTTCGTCTCTGGAGGGGACTTCAGGCCCTGGCTGATCGAGATC 1142
DB	480	GAGCTCTATGGCGCTGACTTCGTGTTCGGGGAGGACTTCAGGCCCTGGCTGATGAGATC 539
QY	1143	AAATTCAGGCCCAACCATGCACCCGCTCCACCGCGTCAAGGCCCGAGCTGTGTCACAGGTG 1202
DB	540	AACGCGAGCCCAACGATGGCAACCTTCCACAGCAGTCACTGCCCGGCTGTGTCTGGCGTG 599
QY	1203	CAGGAGGACAC-----CATCAAGGTTGCCGTGGACCGCAGCTGTGTGACATCGGCA 1251
DB	600	CAAGCTGAAACCTGCGCGTGGTCAATTGACCGGATGCTGGACCGCAACTGTGACACAGGAG 659
QY	1252	ACTTCAGCTCTGTGGAGGCGAGCC 1276
DB	660	CTTTTGAGCTCATCTATAGCAGGC 684

RESULT 14	
CK653583	
LOCUS	
DEFINITION	CK653583 861 bp mRNA linear EST 29-JAN-2004
IMAGE:	AGENCOURT_17673480 NIH_MGC_237 Rattus norvegicus cDNA clone
IMAGE:	7113394 5', mRNA sequence.
ACCESSION	CK653583
VERSION	CK653583.1
KEYWORDS	GI:41389106
SOURCE	EST.
ORGANISM	Rattus norvegicus (Norway rat)
	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.

REFERENCE 1 (bases 1 to 861)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Prepared by: Express Genomics
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
found through the I.M.A.G.E. Consortium/LLNL at:
Plate: LLAM14985 row: e column: 08
High quality sequence stop: 304.
Location/Qualifiers
source
1. .861
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7113394"
/tissue_type="testis, pooled"
/lab_host="DH10B TonA"
/clone_lib="NIH MGC 237"
/note="Organ: Testis; Vector: pExpress-1; Site: 1: EcoRV;
Site 2: NotI; RNA obtained from testis tissue of 8 wk old
animal. Tissues were snap-frozen and kept at -80C before
RNA extraction and purification (Tri-reagent method). cDNA
was primed using oligo-dT primer:
5'-pGACTAGTTGTAGTCGAGCGGCCCT(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb
resulted in an average insert size of 2.4 kb. This primary
library is not normalized (normalized primary library is
NIH_MGC_238) and was constructed by Express Genomics
(Frederick, MD)"

ORIGIN
Query Match 17.2%; Score 316; DB 7; Length 861;
Best Local Similarity 74.2%; Pred. No. 9.4e-59;
Matches 439; Conservative 0; Mismatches 150; Indels 3; Gaps 3;
QY 595 CCGGGGCGGACATAGTGTGCTAGGACCGTGTGGAGAGATCTCTGGAGCTGGCAGCTG 654
DB 12 CCGGGATTGGAGATATTGTGTGATGGACCGTGTGGAGAACATCTCTGGATCTGGTGGCC- 70
QY 655 GAGACCACTCTTTCCAGGGACAAAGTGGGTGTCAGAAATGAGTACATCGAGACCGCC 714
DB 71 CAGACAGGCGAGACCACGAAAGGACAAACAAATGGGTGTCATAGTACATCGAGACCGCGA 130
QY 715 TGCTCATCTGTGACACCAAGTTCGATCAGACAGTGGTTCCTCGTCACGGACTGGAACC 774
DB 131 TGCTCATCTATGACACCAAGTTTGATCAGACAGTGGTTCCTGTTCACAGACTGGAATC 190
QY 775 CCTGACCATCTGGTTCTACAAGGAGTACTTGGCGTTCTCAACTCAGCGCTTCTCCC 834
DB 191 CCTTAACCATCTGGTTCTACAAGAGAGTACTCTCGGTTCTCCAGCGCGCTTCTCCC 250
QY 835 TGACACAGCTGACAGCGCCCTCCCTGTGCAACGACCGCTGCAAGTACCTGAAGA 894
DB 251 TGGACAAACTGACAGCGCTTCCACTGTGTAAACAACTCCGTCGAGGCGGTGGCATAA 310
QY 895 ATGATGTGGGCGCGACCCCTCTGTCGCCGACACAAATGTGGACACACAGCTTCC 954
DB 311 ATGACAAAACGAAGTCCACTGTCTACCTTGGCATTAACATGTGGACCATCACCGCTTCC 370
QY 955 AGGAGTACCTGACGCCAGGCGCGTGGCGCGTGTGGGGCAGCGCTCATCTACCCGTCCA 1014
DB 371 ATGAGTACCTGCATAATAGGGGCGAGGATGCACATGGGGTAGTATCATCTACCCGTCTA 430
QY 1015 TGAAGAGGCCATCGC-CCACGCCATGAAGGTGGGCCCGACGACCATCGTGAGCGCTCGCAAG 1073

Db 431 TGAAGAAGCTCTCACTCAACGTCAATGATGGTGGCCAGATCTTGTGAAGCCCTAAG 490
QY 1074 AACAGCTTTAGCTCTTACGGGCTGACTTCGTCCTTGGAGGAGCTT-CAGGCCCTGGCT 1132
DB 491 ATTATCTTCGATCTCTATGGAGCTGACTTTAATCTGGGCGGTGACTTATAGTCTCTGTGT 550
QY 1133 GATCGAGATCAATTCAGCCGCCACCATGACCGCTCCAGCGGTACCGCC 1184
DB 551 CTTCAAGATCAACTTCTACTCTGCCTTGCAGCTCCTACTTTGTAAAAACC 602

RESULT 15
CD515431 896 bp mRNA linear EST 06-JUN-2003
AGENCOURT 14364765 NIH_MGC.181 Homo sapiens cDNA clone
IMAGE:30407505 5', mRNA sequence.
ACCESSION CD515431
VERSION CD515431.1 GI:314447149
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 896)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM488 row: p column: 10
High quality sequence stop: 678.
Location/Qualifiers
source
1. .896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30407505"
/tissue_type="White Matter"
/dev_stage="Unknown"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/clone_lib="NIH MGC 181"
/note="Vector: PCMV-SPOT6.1; Site 1: NotI; Site 2: EcoRV
(destroyed); Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.42 kb. Library was constructed by
(Invitrogen). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 16.7%; Score 307.6; DB 6; Length 896;
Best Local Similarity 70.0%; Pred. No. 6.6e-57;
Matches 434; Conservative 0; Mismatches 174; Indels 12; Gaps 1;
QY 672 AGGGACAAACAGTGGGTGGTCCAGAGTACATCGAGACCGCGTGTCTATCTGTGACACC 731
DB 25 AAGGACGGCAAGTGGGTGGTGGAGATATTGAGCGGCCCTCTCATCTTTGGCACC 84
QY 732 AAGTTCGACATCAGACAGTGGTTCCTCGTCACGAGCTGGAACCCCTGACCATCTCGTTC 791
DB 85 AAGTTTGACCTCAGACAGTGGTTCCTCGTAACTGACTGGAACCCACTTACCGTGTGGTTC 144
QY 792 TACAAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTCTCCCTGGACAGCTGGACAGC 851


```
Db 145 TACGGACAGCTATATCGCTTTCCACGAGCCCTTCTCCCTGGAAGAACCTGGACAAC 204
Qy 852 GCCATCCACCTGTGCAACACGCGTCCAGAGTACCTGAAGATGATGTGGGCGCGCAGC 911
Db 205 TCAGTGCACCTGTGCAACAACTCCATCCAGAAGCACCTGGAGAACTCATGCCATCGGCAT 264
Qy 912 CCCTGTGCTGCCCGCACACAACATGTGGACAGCACACAGGTTCCAGGAGTACCTGCGAGCGC 971
Db 265 CCAGTGTTCGGCCAGNACACATGTGTCTAGCCAGAGGTTCCAGGCCACCTGCGAGGAG 324
Qy 972 CAGGGCGGTGGCGCGGTGTGGGCGAGCGTCACTTAACCGTCCATGAAGAAGGCCATCGCC 1031
Db 325 ATGGGTGCCCCAAATGCTTGGTCCACCATCATCGTGCCTGGCATGAAGGATGCTGTGATC 384
Qy 1032 CAGCCATGAAGTGGCCAGGACCAAGTGGAGGCTCGCAAGAACAGCTTTGAGCTCTAC 1091
Db 385 CACGCACTTCAGAACCTCCCAAGGACCCGTGCAATGTGGAAGGCCAGCTTTGAGCTCTAT 444
Qy 1092 GGGGCTGACTTTCCTTGGGAGGACTTCAGGCCCTGGCTGATCGAGATCAATTCAGC 1151
Db 445 GGGCTGACTTCGTGTTGCGGGAGGACTTCCAGCCCTGGCTGATTGAGATCAACGCCAGC 504
Qy 1152 CCCACCATGCACCCGTCCACGCCGTGTCACGCCCGCAGCTGTGTGCACAGGTGCGAGGAGC 1211
Db 505 CCCACGATGGCACCCCTCCACAGCAGTCACTGCCCGGCTCTGTGCTGGCGTGCAGCTGAC 564
Qy 1212 ACCATCAAGGTGGCC-----GTGGACCGCAGCTGTGACATCGGCAACTTTGAG 1259
Db 565 ACCCTGCGCGTGTCTATTGACCGGAGGCTGGACCGCAACTGTGACACAGGAGCCTTTGAG 624
Qy 1260 CTCCTGTGGAGGCGAGCGGT 1279
Db 625 CTCATCTATAAGCAGCCCGT 644
```

Search completed: September 25, 2005, 01:55:59
Job time : 7070.48 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 01:51:04 ; Search time 1115.21 Seconds
(without alignments)
9756.464 Million cell updates/sec

Title: US-10-615-659-1

Perfect score: 1838

Sequence: 1 tggagcagccctggggccc.....actctcccaaggcggaattc 1838

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1838	100.0	1838	12	ADJ93357 Human BGS
2	1782.8	97.0	3554	12	ADJ93364 Human tub
3	1678.8	91.3	3465	12	ADJ93363 Human BGS
4	1677.2	91.3	1939	12	ADJ93361 Human BGS
5	1574.8	85.7	1859	12	ADJ93362 Human BGS
6	726	39.5	726	12	ADJ93359 Human BGS
7	501	27.3	101270	12	ADQ17814 Human sof
8	432.8	23.5	2380	6	AAS9894 Polynucle
9	431.6	23.5	2326	4	AAL16735 Human cDN
10	427	23.2	490	12	ACH91699 Human gen
11	421.4	22.9	2538	7	AD873168 Human kid
12	421.4	22.9	2553	13	ACN37881 Tumour-as
13	421.4	22.9	2881	3	AAC77214 Human ORF
14	421.4	22.9	3001	4	RAI58606 Human pol
15	421.4	22.9	3001	5	ADQ98824 DNA encod
16	421.4	22.9	3001	9	ADB48584 Novel hum
17	419.8	22.8	3294	8	ABX34502 Human mdd
18	418	22.7	2848	11	ADM03081 Human cDN
19	413	22.5	2611	3	AAF21812 Human bre
20	347.4	18.9	2412	12	ADH45468 Human mol

21	276	15.0	2469	12	ADH45478 Human mol
22	272.4	14.8	2241	12	ADJ93378 Human BGS
23	231.4	12.6	5282	12	ADQ64426 Novel hum
24	231	12.6	2044	4	AAI60392 Human pol
c	224.6	12.2	3828	13	ADR07582 Full leng
26	213	11.6	2979	4	ABL17657 Drosophil
27	206.8	11.3	2217	10	ACA92443 DNA encod
28	192.4	10.5	1567	8	ACC46592 Human dit
29	190.6	10.4	1958	4	AAL18689 Human cDN
c	190.6	10.4	10468	4	AAK77961 Human imm
31	174.2	9.5	2543	4	ABL17449 Drosophil
32	166	9.0	1728	5	AAS70158 DNA encod
33	164.4	8.9	2250	5	AAS69383 DNA encod
34	163.4	8.9	418	8	ABX52103 Bovine ES
35	147.8	8.0	1085	4	AAF63819 Human sec
36	144.8	7.9	4615	4	ABL17448 Drosophil
c	136.8	7.4	5728	4	ABL17656 Drosophil
38	131.8	7.2	492	9	ACH48041 Human lun
c	130	7.1	525	12	ACH77979 Human gen
40	92.2	5.0	207	6	ABS69387 Novel mur
c	80	4.4	80	12	ADJ93402 Human BGS
42	79.4	4.3	1571	4	ABL14515 Drosophil
43	72.8	4.0	996	10	ADC32299 Human nov
44	72.8	4.0	2259	12	ADQ64293 Novel hum
45	72.8	4.0	2417	10	ADC30440 Human nov

ALIGNMENTS

RESULT 1

ADJ93357

ID ADJ93357 standard; cDNA; 1838 BP.

XX

AC ADJ93357;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human BGS-42 cDNA sequence SeqID1.

XX

KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-Hiv; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.

XX Homo sapiens.

OS

Key Location/Qualifiers

XX CDS

153..1778

/*tag= a

/product= "Human BGS-42 protein"

XX WO2004005487-A2.

XX

XX 15-JAN-2004.

XX

XX 09-JUL-2003; 2003WO-US021605.

XX

XX 09-JUL-2002; 2002US-0394725P.

XX

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

XX Feder JN, Wu, S, Nelson TC;

XX WPI; 2004-099381/10.
DR P-PSDB; ADJ93358.
XX
PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Claim 1; SEQ ID NO 1; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antichratic, antisthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.
XX
SQ Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Query Match 100.0%; Score 1838; DB 12; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGAGCAGCCCTGGGCCCATCTCGGCTATGAGGGCGGGAAGTGGGGCGGCTTGGGGAGCC 60
DB 1 TGGGAGCAGCCCTGGGGCCCATCTCGGCTATGAGGGCGGGAAGTGGGGCGGCTTGGGGAGCC 60

QY 61 TCGTGGCCCTGGGCCCATCTCCAGTCCCAGTCCCAGTCTCGGCTCGGACAGATAGGGCGAGGCT 120
DB 61 TCGTGGCCCTGGGCCCATCTCCAGTCCCAGTCCCAGTCTCGGCTCGGACAGATAGGGCGAGGCT 120

QY 121 GTGCTGTCTTTTCAGAGACTTCCGGCGCACCATGTCATCCAGTCCCTCGGCTCGGACAGATAGGGCGAGG 180
DB 121 GTGCTGTCTTTTCAGAGACTTCCGGCGCACCATGTCATCCAGTCCCTCGGCTCGGACAGATAGGGCGAGG 180

QY 181 TCAGCCACACAGAGCTCCAGCAGGAGCAGCAGAGCAAGCCAGGGACACAGAGGAGGAGG 240
DB 181 TCAGCCACACAGAGCTCCAGCAGGAGCAGCAGAGCAAGCCAGGGACACAGAGGAGGAGG 240

QY 241 CCGGGAGCAGCAGCTTGGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAAAAGCTCAGGG 300
DB 241 CCGGGAGCAGCAGCTTGGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAAAAGCTCAGGG 300

QY 301 GCCTCCGGGGGAGCTTGTGGACATCGCTGCAAGTGTGCCAGGCTTACCTGGGGCAGC 360
DB 301 GCCTCCGGGGGAGCTTGTGGACATCGCTGCAAGTGTGCCAGGCTTACCTGGGGCAGC 360

QY 361 TGGAGCATGAGGACATCGACAGCTGAGCAGATGCCCCTGGAGGAGCTTACCTGAGGGCCAGT 420
DB 361 TGGAGCATGAGGACATCGACAGCTGAGCAGATGCCCCTGGAGGAGCTTACCTGAGGGCCAGT 420

QY 421 GGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATT 480

DB 421 GGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCATCTCCAAATT 480
QY 481 CAAGAAATTTACTTTTCGCAAGTCCAGGCTCTGTAATAGAAATCAGTCTGTGTGAACCCCTC 540
DB 481 CAAGAAATTTACTTTTCGCAAGTCCAGGCTCTGTAATAGAAATCAGTCTGTGTGAACCCCTC 540
QY 541 AGACGGACATTGACCGGGCTCCGGAAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600
DB 541 AGACGGACATTGACCGGGCTCCGGAAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600
QY 601 GCGGAGACATAGTGTGATGACATGACCCGTTGGAGAGAGATCTCGAGCTGGAGCTGCAGACC 660
DB 601 GCGGAGACATAGTGTGATGACATGACCCGTTGGAGAGAGATCTCGAGCTGGAGCTGCAGACC 660
QY 661 ACCCTCTTTCAGGAGCAACAGTGGTGGTCCAGAGTACATCCAGAGCCGCGCTCTCA 720
DB 661 ACCCTCTTTCAGGAGCAACAGTGGTGGTCCAGAGTACATCCAGAGCCGCGCTCTCA 720
QY 721 TCTGTGACACCAAGTTCCGACATCAGACAGTGGTTCCTCTCACGGACTGGAACCCCTGA 780
DB 721 TCTGTGACACCAAGTTCCGACATCAGACAGTGGTTCCTCTCACGGACTGGAACCCCTGA 780
QY 781 CCATCTGGTTCTACAGGAGAGTTACTTGGGTTCCTCAACTCAGCGCTTCTCCCTGGACA 840
DB 781 CCATCTGGTTCTACAGGAGAGTTACTTGGGTTCCTCAACTCAGCGCTTCTCCCTGGACA 840
QY 841 AGCTGACAGAGCCCATCCACCTGTGCAACAACGCGTCCAGAGTACCTGGAAGATGATG 900
DB 841 AGCTGACAGAGCCCATCCACCTGTGCAACAACGCGTCCAGAGTACCTGGAAGATGATG 900
QY 901 TGGGGCGGAGCCCTCTGCTGCCCGCACACAATGTGGACACAGCAGTTCAGGAGT 960
DB 901 TGGGGCGGAGCCCTCTGCTGCCCGCACACAATGTGGACACAGCAGTTCAGGAGT 960
QY 961 ACTGTGAGGCGCAGGCGCTGGGGCGGTGGGGAGCTGATCTTACCCGCTCATGAAGA 1020
DB 961 ACTGTGAGGCGCAGGCGCTGGGGCGGTGGGGAGCTGATCTTACCCGCTCATGAAGA 1020
QY 1021 AGGCCATCGCCACCGCATGAAAGTGGCCAGGACACGCTGGAGCTCCAGAGACAGCT 1080
DB 1021 AGGCCATCGCCACCGCATGAAAGTGGCCAGGACACGCTGGAGCTCCAGAGACAGCT 1080
QY 1081 TTGAGCTCTACCGGGCTGACTTGGTCTTGGAGGAGTTCAGGCTTCGCTGATCGAGA 1140
DB 1081 TTGAGCTCTACCGGGCTGACTTGGTCTTGGAGGAGTTCAGGCTTCGCTGATCGAGA 1140
QY 1141 TCAATTCCAGCCCAACATGCAACCGTTCACCGGGTCAAGGCCACAGCTGTGTGACAGG 1200
DB 1141 TCAATTCCAGCCCAACATGCAACCGTTCACCGGGTCAAGGCCACAGCTGTGTGACAGG 1200
QY 1201 TGCAGGAGGACACCATCAAGTGGCCGTGACCGAGCTGTGACATCGGCAACTTCGAGC 1260
DB 1201 TGCAGGAGGACACCATCAAGTGGCCGTGACCGAGCTGTGACATCGGCAACTTCGAGC 1260
QY 1261 TCGTGTGAGGAGCAGCGGTGGTGGCGCCCAATTCAGCGGCTCCGACCTTCGCTGG 1320
DB 1261 TCGTGTGAGGAGCAGCGGTGGTGGCGCCCAATTCAGCGGCTCCGACCTTCGCTGG 1320
QY 1321 CCGGGCTCTAGTGTGAGGAGCAGGAGTGTGCGCTTGTGCAACTTCGAGGCTTCAGGCT 1380
DB 1321 CCGGGCTCTAGTGTGAGGAGCAGGAGTGTGCGCTTGTGCAACTTCGAGGCTTCAGGCT 1380
QY 1381 CCGGCTCTGCTTGTGAGCGCGAGCGCTGAGGACCGGGCCCTTCGGCCATTCGAGACC 1440
DB 1381 CCGGCTCTGCTTGTGAGCGCGAGCGCTGAGGACCGGGCCCTTCGGCCATTCGAGACC 1440
QY 1441 CTGCCAGGAGCCCACTCAGAGCTCTCAGCGGACTTGGGACTTGAAGAGAGAAG 1500
DB 1441 CTGCCAGGAGCCCACTCAGAGCTCTCAGCGGACTTGGGACTTGAAGAGAGAAG 1500
QY 1501 GGCTCCCTCTGGCTTGTGGCACCTTAAAGGGGGCAGCCGAGAGCTGGAGCCGAC 1560

Db 1501 GGCTCCCGCTGGCTTGTGSCACCCCTTAAGGGGGGCGAGCGGTGGAGCCGCAC 1560

Qy 1561 AGCCACCCGACCAAGAGCTCTGGGAGAGTGGAGCTCCCGGCTGCCCTGTGCCACG 1620

Db 1561 AGCCACCCGACCAAGAGCTCTGGGAGAGTGGAGCTCCCGGCTGCCCTGTGCCACG 1620

Qy 1621 TGGACAGTCAAGCCCAACACCGGTGTCCCGTAGCCGACCGCCGCAAAAGCTGGGATC 1680

Db 1621 TGGACAGTCAAGCCCAACACCGGTGTCCCGTAGCCGACCGCCGCAAAAGCTGGGATC 1680

Qy 1681 CAACACAGCTAAATTCGCGACCCGCTGGAGCCGTGTGCTGCGGGGCTGAAGACAGCAGG 1740

Db 1681 CAACACAGCTAAATTCGCGACCCGCTGGAGCCGTGTGCTGCGGGGCTGAAGACAGCAGG 1740

Qy 1741 GCGGCTGCTCCCGCCCGCGAGGAAAGTTATGACAGCTCAGATTCCTGACGA 1800

Db 1741 GCGGCTGCTCCCGCCCGCGAGGAAAGTTATGACAGCTCAGATTCCTGACGA 1800

Qy 1801 GGAGTACAGGTTGACGACCTCTCCCAAGGCGGAATTC 1838

Db 1801 GGAGTACAGGTTGACGACCTCTCCCAAGGCGGAATTC 1838

RESULT 2

ID ADJ93364 standard; DNA; 3554 BP.

XX ADJ93364;

XX AC ADJ93364;

XX AC ADJ93364;

DT 06-MAY-2004 (first entry)

XX Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;

XX BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;

XX neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;

XX osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;

XX anti-HIV; antibacterial; immunosuppressive; antiseborrheic;

XX dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;

XX tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;

XX testicular disorder; testicular cancer; pulmonary disorder; lung cancer;

XX gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;

XX brain cancer; liver cancer; proliferative condition; testis; lung;

XX small intestine; brain; lymph tissue; infertility; Cushing's syndrome;

XX emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;

XX Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;

XX sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.

XX Homo sapiens.

OS Synthetic.

XX

Key Location/Qualifiers

FT 1695..3320

FT /*tag= a

FT /product= "Human tubulin tyrosine ligase protein"

XX

PN W02004005487-A2.

XX

PD 15-JAN-2004.

XX

PP 09-JUL-2003; 2003WO-US021605.

XX

PP 09-JUL-2002; 2002US-0394725P.

XX

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

PI Feder JN, Wu S, Nelson TC;

XX

XX WPI; 2004-099381/10.

DR P-PSDB; ADJ93365.

XX

PT New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,

PT useful for preventing, treating or ameliorating a medical condition, e.g.

PT aberrant cellular proliferation, reproductive disorders or testicular

PT disorders.

XX

XX Example 4; SEQ ID NO 12; 343bp; English.

XX

CC This invention relates to a novel testis-specific tubulin tyrosine-ligase

CC -like polypeptide, designated the BGS-42 polypeptide. The invention may

CC be useful for the development of compounds with a cytostatic, respiratory

CC -gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,

CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,

CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,

CC immunosuppressive, antiseborrheic or dermatological activity acting as

CC tyrosine ligase modulators. In addition, the disclosed sequences may be

CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be

CC used for diagnosing a pathological condition or a susceptibility to a

CC pathological condition in a subject, and for preventing, treating or

CC ameliorating a medical condition, such as a disorder related to aberrant

CC tubulin ligase activity, a disorder related to aberrant tubulin-

CC carboxypeptidase activity, aberrant cellular proliferation, reproductive

CC disorders, testicular disorders, testicular cancer, pulmonary disorders,

CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,

CC neural disorders, brain cancer, liver cancer, or proliferative condition

CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42

CC polypeptide, polynucleotide, or their modulators are also useful for

CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's

CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS

CC -42 polypeptide can be used as a preventive agent for immunological

CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's

CC disease or scleroderma. The antibodies may be used to purify, detect and

CC target the BGS-42 polypeptides. The present sequence is that of the

CC tubulin tyrosine ligase protein consensus gene sequence which was used in

CC the exemplification of the invention.

XX

Seq Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;

Query Match 97.0%; Score 1782.8; DB 12; Length 3554;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1824; Conservative 0; Mismatches 2; Indels 30; Gaps 1;

Qy 1 TGGGAGCAGGCGCTGGGCGCCCATCGCTATGAGGCGGGAAGTGGGCGGTGGGAGCC 60

Db 1513 TGGGAGCAGGCGCTGGGCGCCCATCGCTATGAGGCGGGAAGTGGGCGGTGGGAGCC 1572

Qy 61 TCGTGGCGCTGGCGCCCATCTCCAGTCCCAGTCTCGGCTCGGACATAGGGCGAGCT 120

Db 1573 TCGTGGCGCTGGCGCCCATCTCCAGTCCCAGTCTCGGCTCGGACATAGGGCGAGCT 1632

Qy 121 GTCTGTCTCTTTCA-----GAAGACTTCCGGCGCA 150

Db 1633 GTCTGTCTCTTTTTCAGAGTACGCGGTGGGGAAGAGGTTCTTGGGAAGACTTCCGGCGCA 1692

Qy 151 CCATGGCATCCAGCATCTCAAGTGGGTGTGTCAGCCACAGAGCTGCAGGAGCAGCA 210

Db 1693 CCATGGCATCCAGCATCTCAAGTGGGTGTGTCAGCCACAGAGCTGCAGGAGCAGCA 1752

Qy 211 GAAGCAAGCCCGGACCCAGAGGAGAGCGCCGGGAGCAGGACCTTAGCAGCAGGCAAG 270

Db 1753 GAAGCAAGCCCGGAGCCAGAGGAGAGCGCCGGGAGCAGGACCTTAGCAGCAGGCAAG 1812

Qy 271 ATGCTGAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGGT 330

Db 1813 ATGCTGAATGCTGAGGCAAGCTCAGGGGCTCCCGGGGAGCTTGTGACATCGGT 1872

Qy 331 GCAAGGTGTGCCAGGCGCTTCTGGGGCAGCTGGAGCATGAGGACATCGACACGTGACAG 390

Db 1873 GCAAGGTGTGCCAGGCGCTTCTGGGGCAGCTGGAGCATGAGGACATCGACACGTGACAG 1932

Qy 391 ATCCGTGGAGGACCTCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 450

Db 1933 ATCCGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1992

Qy 451 TCGTTCATGGCGATGCTTTCATCTCCCAATTCAGAAATTAATTTTCGACGTGCAGGCTC 510

Db 1993 TCCTTCATGCGCATGCTTTTCATCTCCAATTCAGAAATTAATTTTCGACAGTCCAGGCTC 2052

QY 511 TGTGTAATAAATCACTCTGTGAACCTTCAGACGGACATTAGACGGCTCCGGAACATCT 570

Db 2053 TGTGTAATAGANTCAGTCTGTGAACCTTCAGACGGACATTAGACGGCTCCGGAACATCT 2112

QY 571 GGATTATAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGG 630

Db 2113 GGATTATAAGCCCGCGCCCAAGTCCCGGGCCGAGACATAGTGTGCATGGACCGTGTGG 2172

QY 631 AGGAGATCTGGAGCTGGAGCTGGAGACACACCTCTTTCCAGGGACAACAAGTGGGTGG 690

Db 2173 AGGAGATCTGGAGCTGGAGCTGGAGACACACCTCTTTCCAGGGACAACAAGTGGGTGG 2232

QY 691 TCCAGAAGTACATCGAGACGCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT 750

Db 2233 TCCAGAAGTACATCGAGACGCCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGT 2292

QY 751 GGTTCCTCGTCAAGGACTGGAACCCCTGACCAATCTGGTTTCTAAGAGAGAGTTACTTGC 810

Db 2293 GGTTCCTCGTCAAGGACTGGAACCCCTGACCAATCTGGTTTCTAAGAGAGAGTTACTTGC 2352

QY 811 GGTTCCTCAACTAGCGCTTCTCCTCGACAGCTGACAGCGCCATCCACTGTGCAACA 870

Db 2353 GGTTCCTCAACTAGCGCTTCTCCTCGACAGCTGACAGCGCCATCCACTGTGCAACA 2412

QY 871 ACGCCGTTCAGAAAGTACCTGAAGAAATGATGTGGCCGCGAGCCCCCTGCTGCCGCGACACA 930

Db 2413 ACGCCGTTCAGAAAGTACCTGAAGAAATGATGTGGCCGCGAGCCCCCTGCTGCCGCGACACA 2472

QY 931 ACATGTGGACACAGCACCAGGTTCCAGAGGTACCTGACAGCGCCAGGGCCGCTGGCGCGTGT 990

Db 2473 ACATGTGGACACAGCACCAGGTTCCAGAGGTACCTGACAGCGCCAGGGCCGCTGGCGCGTGT 2532

QY 991 GGGGAGCTGATCTATACCGTTCATGAAGAGCCATGGCCACGCCATGAAGTGGCC 1050

Db 2533 GGGGAGCTGATCTATACCGTTCATGAAGAGCCATGGCCACGCCATGAAGTGGCC 2592

QY 1051 AGGACACGTGGAGCTTCGCAAGAACAGCTTTCAGCTCTACGGGGCTGACTTCGTCCTTG 1110

Db 2593 AGGACACGTGGAGCTTCGCAAGAACAGCTTTCAGCTCTACGGGGCTGACTTCGTCCTTG 2652

QY 1111 GGAGGACTTCAGGCCCTGCTGATCGAGATCAATTCAGCCCCACCATGCAACCCGTCCA 1170

Db 2653 GGAGGACTTCAGGCCCTGCTGATCGAGATCAATTCAGCCCCACCATGCAACCCGTCCA 2712

QY 1171 CGCCGCTCAGGCCCGCAGCTGTGTGCACAGTGCAGAGACACCATCAAGTGGCCGTGG 1230

Db 2713 CGCCGCTCAGGCCCGCAGCTGTGTGCACAGTGCAGAGACACCATCAAGTGGCCGTGG 2772

QY 1231 ACCGAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGACGCGGTGGTTGAGCCGC 1290

Db 2773 ACCGAGCTGTGACATCGGCAACTTCGAGCTCTGTGGAGGACGCGGTGGTTGAGCCGC 2832

QY 1291 CCCATTACAGCGGTCCGACCTCTGCTGCGCGGCGTCAAGTGTGAGGAGAGCCAGGAGGC 1350

Db 2833 CCCATTACAGCGGTCCGACCTCTGCTGCGCGGCGTCAAGTGTGAGGAGAGCCAGGAGGC 2892

QY 1351 AGTGTGTCCTGCTGCAACCTCAAGGCTTCGGCTCGCTGTTGGACGCGCAGCCCTGA 1410

Db 2893 AGTGTGTCCTGCTGCAACCTCAAGGCTTCGGCTCGCTGTTGGACGCGCAGCCCTGA 2952

QY 1411 AGGACAGGGGCCCTCGGCCATGCCAGACCTCTCCAGGGACCCCATCAACAGCTCTCC 1470

Db 2953 AGGACAGGGGCCCTCGGCCATGCCAGACCTCTCCAGGGACCCCATCAACAGCTCTCC 3012

QY 1471 AGCGGACTTGGGACTGAAGAGAGAGAGGGGCTCCCTCGGCTTGTGCTGGCACCCCTTAA 1530

Db 3013 AGCGGACTTGGGACTGAAGAGAGAGAGGGGCTCCCTCGGCTTGTGCTGGCACCCCTTAA 3072

QY 1531 GGGGGGACCGGAGAGCGGTGGAGCGGCGACAGCCACCGCACCAAGTGTGGGAAGG 1590

Db 3073 GGGGGGACCGGAGAGCGGTGGAGCGGCGCACAGCCACCGCACCAAGTGTGGGAAGG 3132

QY 1591 TGGAGCTCCCGGCTGCCCCCTGTGCCACGTCGACAGTCAGGCCCAACACACCGGTGTCC 1650

Db 3133 TGGAGCTCCCGGCTGCCCCCTGTGCCACGTCGACAGTCAGGCCCAACACACCGGTGTCC 3192

QY 1651 CCGTAGCCACGAGCCGCCAAAAGCTGGGATCCAAAACAGCTAAATGCGCAACCCGCTGGAGC 1710

Db 3193 CCGTAGCCACGAGCCGCCAAAAGCTGGGATCCAAAACAGCTAAATGAGCACCCGCTGGAGC 3252

QY 1711 CTGTGCTGGGGGCTTGAAGACAGACAGAGCGCGCTGCTGCCGCCCGGAGGAAAAG 1770

Db 3253 CTGTGCTGGGGGCTTGAAGACAGACAGAGCGCGCTGCTGCCGCCCGGAGGAAAAG 3312

QY 1771 GTTCATGACAGGCTCAGATTCTCTGCAGCAGGAGTACAGGTTGCAGCCACTCTCC 1826

Db 3313 GTTCATGACAGGCTCAGATTCTCTGCAGCAGGAGTACAGGTTGCAGCCACTCTCC 3368

RESULT 3

ADJ93363

ID ADJ93363 standard; DNA; 3465 BP.

XX

AC ADJ93363;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human BGS-42 protein-related DNA clone C SeqID11.

XX

KW testis-specific tubulin tyrosine-ligase-like polypeptide;

KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;

KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;

KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;

KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;

KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;

KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;

KW testicular disorder; testicular cancer; pulmonary cancer; lung cancer;

KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;

KW brain cancer; liver cancer; proliferative condition; testis; lung;

KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;

KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;

KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;

KW sepsis; acne; Sjogren's disease; scleroderma; human; db.

XX

OS Homo sapiens.

XX

PN WO2004005487-A2.

XX

PD 15-JAN-2004.

XX

PF 09-JUL-2003; 2003WO-US021605.

XX

PR 09-JUL-2002; 2002US-0394725P.

XX

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

PI Feder JN, Wu S, Nelson TC;

XX

DR WPI; 2004-099381/10.

XX

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,

PT useful for preventing, treating or ameliorating a medical condition, e.g.

PT aberrant cellular proliferation, reproductive disorders or testicular disorders.

PT

XX

PS Example 4; SEQ ID NO 11; 343pp; English.

XX

CC This invention relates to a novel testis-specific tubulin tyrosine-ligase

CC -like polypeptide, designated the BGS-42 polypeptide. The invention may

CC be useful for the development of compounds with a cytostatic, respiratory

CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,

CC antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,

CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,

CC immunosuppressive, antiseborrheic or dermatological activity acting as

tyrosine ligase modulators. In addition, the disclosed sequences may be useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of a DNA clone sequence which is related to the invention.

XX Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Query Match 91.3%; Score 1678.8; DB 12; Length 3465;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1757; Conservative 0; Mismatches 2; Indels 67; Gaps 1;

Qy	1	TGGAGCAGCCCTGGGCCCCATGGCTATGAGGGCGGGAAGTGGGGGGGTGGGAGCC	60
Db	1526	TGGAGCAGCCCTGGGCCCCATGGCTATGAGGGCGGGAAGTGGGGGGGTGGGAGCC	1585
Qy	61	TCCGTGCCCTGGCCCCATCTCCAGTCCCAGTCCCTGGCTCGACATAGGCGGAGGCT	120
Db	1586	TCCGTGCCCTGGCCCCATCTCCAGTCCCAGTCCCTGGCTCGACATAGGCGGAGGCT	1645
Qy	121	GTGCTGTCTTTTTCAGAGACTTCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG	180
Db	1646	GTGCTGTCTTTTTCAGAGACTTCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG	1705
Qy	181	TCAGCCACAGAGCTGACAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	240
Db	1706	TCAGCCACAGAGCTGACAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1765
Qy	241	CCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	300
Db	1766	CCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	1825
Qy	301	GCCTCCCGGGGAGCTTGTGGACATCGCTGCAAGGTGTCCAGGCGCTACTGGGGCAGC	360
Db	1826	GCCTCCCGGGGAGCTTGTGGACATCGCTGCAAGGTGTCCAGGCGCTACTGGGGCAGC	1885
Qy	361	TGGAGCATGAGGACATGACACGTGACAGATGCGGTGAGAGAGCTTACCTGGGCGAGT	420
Db	1886	TGGAGCATGAGGACATGACACGTGACAGATGCGGTGAGAGAGCTTACCTGGGCGAGT	1945
Qy	421	GGGAGGACCTGACCCAGCAGTACTCTCCCTGTTCAATGGCGATGTTTCATCTCCAAAT	480
Db	1946	GGGAGGACCTGACCCAGCAGTACTCTCCCTGTTCAATGGCGATGTTTCATCTCCAAAT	2005
Qy	481	CAAGAAATTAATTTTTCAGTGGCAGGCTCTGCTGAATAGAATCAGCTCTGTGAACCTTC	540
Db	2006	CAAGAAATTAATTTTTCAGTGGCAGGCTCTGCTGAATAGAATCAGCTCTGTGAACCTTC	2065
Qy	541	AGACGACATGAGCGGCTCCGAAACATCTGGATTTAAAGCCCGGCGCAAGTCCCGGG	600
Db	2066	AGACGACATGAGCGGCTCCGAAACATCTGGATTTAAAGCCCGGCGCAAGTCCCGGG	2125
Qy	601	GCCGAGACATGATGTGCATGGAACCGTGTGGAGAGATCTGGAGCTGGCAGTGCAGACC	660
Db	2126	GCCG-----	2129
Qy	661	ACCTCTTTTCCAGGGACAAAGTGGGTGGTCCAGAAATACATCGAGAGCCCGCTGCTCA	720
Db	2130	-----AGGACACACAGTGGGTGGTCCAGAAATACATCGAGAGCCCGCTGCTCA	2178

QY 1801 GGAGTACAGGTTCCAGCCACTCTCCC 1826
 Db |||||||||||||||||||||||||||||
 3259 GGAGTACAGGTTCCAGCCACTCTCCC 3284

RESULT 4

ADJ93361

ID ADJ93361 standard; DNA; 1939 BP.

XX

AC ADJ93361;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human BGS-42 protein-related DNA clone A SeqID9.

XX

KW testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nontropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.

XX Homo sapiens.

XX

OS WO2004005487-A2.

XX

PN 15-JAN-2004.

XX

PD 09-JUL-2003; 2003WO-US021605.

XX

PF 09-JUL-2002; 2002US-0394725P.

XX

PR (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

PA Feder JN, Wu S, Nelson TC;

XX

PI WPI; 2004-099381/10.

XX

DR New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 XX useful for preventing, treating or ameliorating a medical condition, e.g.
 XX aberrant cellular proliferation, reproductive disorders or testicular
 XX disorders.

PS Example 4; SEQ ID NO 9; 343pp; English.

XX

PS This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytostatic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nontropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's

CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of a DNA
 CC clone sequence which is related to the invention.

XX

SQ Sequence 1939 BP; 421 A; 592 C; 589 G; 337 T; 0 U; 0 Other;

XX

Query Match 91.3%; Score 1677.2; DB 12; Length 1939;

XX

Best Local Similarity 96.2%; Pred. No. 0;

XX

Matches 1756; Conservative 0; Mismatches 3; Indels 67; Gaps 1;

XX

QY 1 TGGGAGCAGGCTGGGCCCCCATCGGCTATGAGGGCGGAAGTGGGGCGGTTTGGGAGCC 60

Db

1 TGGGAGCAGGCTGGGCCCCCATCGGCTATGAGGGCGGAAGTGGGGCGGTTTGGGAGCC 60

QY

61 TCGGTGGCCCTGGCCCCCATCTCCAGTCCCACTTCCGCTCGGACAGATAGGCGGAGGCT 120

Db

61 TCGGTGGCCCTGGCCCCCATCTCCAGTCCCACTTCCGCTCGGACAGATAGGCGGAGGCT 120

QY

121 GTGCTGTCTTTTCAGAGAGACTTCCCGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG 180

Db

121 GTGCTGTCTTTTCAGAGAGACTTCCCGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG 180

QY

181 TCAGCCACCCAGAGCTGCGAGGAGCAGCAGAAGCAAGCCAGGAGCCAGAGGGAGG 240

Db

181 TCAGCCACCCAGAGCTGCGAGGAGCAGCAGAAGCAAGCCAGGAGCCAGAGGGAGG 240

QY

241 CCGGAGCAGCGACCTGAGCAGCAGCAAGATGCTGAAAATCTGAGCAAAAGCTCAGGG 300

Db

241 CCGGAGCAGCGACCTGAGCAGCAGCAAGATGCTGAAAATCTGAGCAAAAGCTCAGGG 300

QY

301 GCCTCCCGGGGCGAGCTTGTGGACATCGGCTGCAAGTGTCCAGGCTACCTGGGGCAGC 360

Db

301 GCCTCCCGGGGCGAGCTTGTGGACATCGGCTGCAAGTGTCCAGGCTACCTGGGGCAGC 360

QY

361 TGGAGCATGAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTAGGGCGAGT 420

Db

361 TGGAGCATGAGGACATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTAGGGCGAGT 420

QY

421 GGGAGGACCTGACCCAGCAGTACTACTCCCTGTTTCATGGCGATGCTTCAATCTCAATT 480

Db

421 GGGAGGACCTGACCCAGCAGTACTACTCCCTGTTTCATGGCGATGCTTCAATCTCAATT 480

QY

481 CAAGAAATTTACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCAGTCTGTGAACCTTC 540

Db

481 CAAGAAATTTACTTTTCGAGTGCAGGCTCTGCTGAATAGAAATCAGTCTGTGAACCTTC 540

QY

541 AGACGGACATTGACGGGCTCCGGAAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600

Db

541 AGACGGACATTGACGGGCTCCGGAAACATCTGGATTATAAAGCCCGCGGCAAGTCCCGGG 600

QY

601 GCGGAGACATAGTGTGCATGGACCTGTGGAGGAGATCTGGAGCTGGCAGTGCAGACC 660

Db

601 GCGGAGACATAGTGTGCATGGACCTGTGGAGGAGATCTGGAGCTGGCAGTGCAGACC 660

QY

661 ACCCTCTTTTCAGGAGCAACAAAGTGGGTGGTCCAGAAATACATCGAGAGCCGCTGCTCA 720

Db

605 -----AGGAGCAACAAAGTGGGTGGTCCAGAAATACATCGAGAGCCGCTGCTCA 653

QY

721 TCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTACGGAGCTGGAAACCCCTGA 780

Db

654 TCTGTGACACCAAGTTCGACATCAGACAGTGGTTCCTCGTACGGAGCTGGAAACCCCTGA 713

QY

781 CCATCTGTTTTCAGAGGAGGATTAATGCGGTTCTCACTCAGCGCTTCTCCCTGGACA 840

Db

714 CCATCTGTTTTCAGAGGAGGATTAATGCGGTTCTCACTCAGCGCTTCTCCCTGGACA 773

QY

841 AGCTGGACAGCGCCATCCACCTGTGCAACCAAGCCGTCAGAAAGTACCTGAAGAATGATG 900

Db

774 AGCTGGACAGCGCCATCCACCTGTGCAACCAAGCCGTCAGAAAGTACCTGAAGAATGATG 833


```
QY 901 TGGGCGCAGCCCTCTGTCGCCGACACACCAACATGTGACCGACAGCACCAGGTTCCAGGAGT 960
D 834 TGGGCGCAGCCCTCTGTCGCCGACACACCAACATGTGACCGACAGCACCAGGTTCCAGGAGT 893
QY 961 ACCTGACGCCAGGGCGGTGGGCGCGTGTGGGCGAGCGTCATCTACCCCTCCATGAAGA 1020
D 894 ACCTGACGCCAGGGCGGTGGGCGCGTGTGGGCGAGCGTCATCTACCCCTCCATGAAGA 953
QY 1021 AGGCGATCGCCAGCCCATGAAGTGGCCACAGACACAGTGGAGCTTCGCAAGAACAGCT 1080
D 954 AGGCGATCGCCAGCCCATGAAGTGGCCACAGACACAGTGGAGCTTCGCAAGAACAGCT 1013
QY 1081 TTGAGCTCTACGGGCGTACTTGTCTCTTGGAGGAGCTTCAGGCGCTGCTGATCGAGA 1140
D 1014 TTGAGCTCTACGGGCGTACTTGTCTCTTGGAGGAGCTTCAGGCGCTGCTGATCGAGA 1073
QY 1141 TCAATTCAGCCCCACCATGACACCCGTCACGCGCGTTCAGGCGCCAGCTGTGTGCACAGG 1200
D 1074 TCAATTCAGCCCCACCATGACACCCGTCACGCGCGTTCAGGCGCCAGCTGTGTGCACAGG 1133
QY 1201 TGCAGGAGGACCATCAAGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGC 1260
D 1134 TGCAGGAGGACCATCAAGTGGCCCGTGGACCGCAGCTGTGACATCGGCAACTTCGAGC 1193
QY 1261 TCCTGTGGAGGACCGGTGTGAGCGCCGCCCATTTTCAGGGGTCGACCTCTGGTGG 1320
D 1194 TCCTGTGGAGGACCGGTGTGAGCGCCGCCCATTTTCAGGGGTCGACCTCTGGTGG 1253
QY 1321 CGGCGCTCAGTGTGAGGAGGACGAGGAGGAGTGTGCTGCCCGTTCGCAACTCAAGGCGCT 1380
D 1254 CGGCGCTCAGTGTGAGGAGGACGAGGAGGAGTGTGCTGCCCGTTCGCAACTCAAGGCGCT 1313
QY 1381 CGGCGCTCAGTGTGAGGAGGACGAGGAGGAGTGTGAGGAGGAGTGTGAGGAGGAGGAGG 1440
D 1314 CGGCGCTCAGTGTGAGGAGGACGAGGAGGAGTGTGAGGAGGAGTGTGAGGAGGAGGAGG 1373
QY 1441 CTGCGCAGGACCCCATCATCAGCTCTCCAGGGGAGCTTGGGAGTGTGAGGAGGAGGAGG 1500
D 1374 CTGCGCAGGACCCCATCATCAGCTCTCCAGGGGAGCTTGGGAGTGTGAGGAGGAGGAGG 1433
QY 1501 GGTCTCCCTTGGCTTGTGCTGGCACCCTTAAAGGGGGGAGCGGAGGAGCGGTGGAGCGCGCAC 1560
D 1434 GGTCTCCCTTGGCTTGTGCTGGCACCCTTAAAGGGGGGAGCGGAGGAGCGGTGGAGCGCGCAC 1493
QY 1561 AGCCACCCCGACAAAGCTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAG 1620
D 1494 AGCCACCCCGACAAAGCTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAG 1553
QY 1621 TGGACAGTCAAGCCCAACACCGGTGTCCCGTGGCCAGCCCGCCCAAGCTGGGATC 1680
D 1554 TGGACAGTCAAGCCCAACACCGGTGTCCCGTGGCCAGCCCGCCCAAGCTGGGATC 1613
QY 1681 CAAACACAGCTAAATGCGACCCGCTGGAGGCTGTGCTGGCGGCGCTGAACAGCAGCAGG 1740
D 1614 CAAACACAGCTAAATGCGACCCGCTGGAGGCTGTGCTGGCGGCGCTGAACAGCAGCAGG 1673
QY 1741 GCGCGTGTCTCCCGCCCGGAGGAGGAGGTTATGACAGCTGTGATCTCTGAGCA 1800
D 1674 GCGCGTGTCTCCCGCCCGGAGGAGGAGGTTATGACAGCTGTGATCTCTGAGCA 1733
QY 1801 GGAGTACAGTTGCGACCACTCTCCC 1826
D 1734 GGAGTACAGTTGCGACCACTCTCCC 1759
```

RESULT 5

ADJ93362

ID ADJ93362 standard; DNA; 1859 BP.

XX

AC ADJ93362;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human BGS-42 protein-related DNA clone B SeqID10.

XX testis-specific tubulin tyrosine-ligase-like polypeptide;

KW BGS-42 polypeptide; cytosolic; respiratory; anti-inflammatory; anabolic; hypertensive;

KW neuroprotective; endocrine; anti-inflammatory; anabolic; hypertensive;

KW osteopathic; nontoxic; antiparkinsonian; antiarthritic; antiasthmatic;

KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;

KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;

KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;

KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;

KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;

KW brain cancer; liver cancer; proliferative condition; testis; lung;

KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;

KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;

KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;

KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.

XX Homo sapiens.

OS WO2004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

DR

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,

PT useful for preventing, treating or ameliorating a medical condition, e.g.

PT aberrant cellular proliferation, reproductive disorders or testicular

PT disorders.

XX Example 4; SEQ ID NO 10; 343pp; English.

XX This invention relates to a novel testis-specific tubulin tyrosine-ligase

CC -like polypeptide, designated the BGS-42 polypeptide. The invention may

CC be useful for the development of compounds with a cytostatic, respiratory

CC -gen, gastrointestinal-gen, neuroprotective, endocrine-gen,

CC anti-inflammatory, anabolic, hypertensive, osteopathic, nontoxic,

CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,

CC immunosuppressive, antiseborrheic or dermatological activity acting as

CC tyrosine ligase modulators. In addition, the disclosed sequences may be

CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be

CC used for diagnosing a pathological condition or a susceptibility to a

CC pathological condition in a subject, and for preventing, treating or

CC ameliorating a medical condition, such as a disorder related to aberrant

CC tubulin ligase activity, a disorder related to aberrant tubulin-

CC carboxypeptidase activity, aberrant cellular proliferation, reproductive

CC disorders, testicular disorders, testicular cancer, pulmonary disorders,

CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,

CC neural disorders, brain cancer, liver cancer, or proliferative condition

CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42

CC polypeptide, polynucleotide, or their modulators are also useful for

CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's

CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS

CC -42 polypeptide can be used as a preventive agent for immunological

CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's

CC disease or scleroderma. The antibodies may be used to purify, detect and

CC target the BGS-42 polypeptides. The present sequence is that of a DNA

CC clone sequence which is related to the invention.

XX

XX Sequence 1859 BP; 418 A; 563 C; 562 G; 316 T; 0 U; 0 Other;

Query Match 85.7%; Score 1574.8; DB 12; Length 1859;

Best Local Similarity 97.1%; Pred. No. 0;

Matches 1643; Conservative 0; Mismatches 2; Indels 47; Gaps 2;

Qy	135	GAAGACTTTCGGCGCGACAATGGGCATCCAGTCAATCTCAAGTGGGTGGTTCAGCCACGAGC	194
Db	29	GAAGACTTTCGGCGCGACAATGGGCATCCAGTCAATCTCAAGTGGGTGGTTCAGCCACGAGC	88
Qy	195	TGCAGCAGGAGCAGCAGAACGCCAGCGGACAGAGGGAGGAGGCCGGGAGCAGCGAC	254
Db	89	TGCAGCAGGAGCAGCAGAACGCCAGCGGACAGAGGGAGGAGGCCGGGAGCAGCGAC	148
Qy	255	CTGAGCAGCAGGCAAGATGCTGAAAAATGCTGAGGCAAAAGCTTCAGGGGCGCTCCCGGGCGAG	314
Db	149	CTGAGCAGCAGGCAAGATGCTGAAAAATGCTGAGGCAAAAGCTTCAGGGGCGCTCCCGGGCGAG	208
Qy	315	CTTGTGGACATTCGGTGCAGAGTGTGCAGGCTTACTCTGGGGCAGCTGGGAGCATGAGGAC	374
Db	209	CTTGTGGACATTCGGTGCAGAGTGTGCAGGCTTACTCTGGGGCAGCTGGGAGCATGAGGAC	268
Qy	375	ATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCTGACC	434
Db	269	ATCGACACGTCAGCAGATGCCGTGGAGGACCTCACTGAGGCCGAGTGGGAGGACCTGACC	328
Qy	435	CAGCAGTACTACTCCCTCGTTTCATGGCGATGCTTTTCACTCATCAAGAAATTTACTTT	494
Db	329	CAGCAGTACTACTCCCTCGTTTCA-----	351
Qy	495	TCCGAGTGCAGGCTCTGTGTAATAGAAATCACGTCTGTGAACCTTCAGACGGACATTTGAC	554
Db	352	-----TGCCAGGCTCTGTGTAATAGAAATCACGTCTGTGAACCTTCAGACGGACATTTGAC	405
Qy	555	GGGCTCCGGAACATCTCGATTATAAAGCCCGCGGCCAAGTCCCGGGGCCAGACATAGTG	614
Db	406	GGGCTCCGGAACATCTCGATTATAAAGCCCGCGGCCAAGTCCCGGGGCCAGACATAGTG	465
Qy	615	TGCATGACACGCTGTGGAGGAGATCCTCGAGCTGGCAGCTGCAGACCACTCTTTTCCAGG	674
Db	466	TGCATGACACGCTGTGGAGGAGATCCTCGAGCTGGCAGCTGCAGACCACTCTTTTCCAGG	525
Qy	675	GACAAAGTGGGTGGTCCAGAAAGTACATCGAGACGCCGTGCTCATCTGTGACACCAAG	734
Db	526	GACAAAGTGGGTGGTCCAGAAAGTACATCGAGACGCCGTGCTCATCTGTGACACCAAG	585
Qy	735	TTTCGACATCAGAAGTGGTTCCTCGTCAAGGACTGAGAACCCCTTGACCATCTGGTTCTAC	794
Db	586	TTTCGACATCAGAAGTGGTTCCTCGTCAAGGACTGAGAACCCCTTGACCATCTGGTTCTAC	645
Qy	795	AAGGAGAGTTACTTTGCGGTTCTCAAATCAGCGCTTCTCCCTGGACACAGCTGGACGGCC	854
Db	646	AAGGAGAGTTACTTTGCGGTTCTCAAATCAGCGCTTCTCCCTGGACACAGCTGGACGGCC	705
Qy	855	ATCCACCTGTGCAACAAACGCCGTCCAGAAAGTACCTGAAGAAATGATGTGGGCCGACGCC	914
Db	706	ATCCACCTGTGCAACAAACGCCGTCCAGAAAGTACCTGAAGAAATGATGTGGGCCGACGCC	765
Qy	915	CTGCTGCCCGGACACAAATGTGTGAACAGCAACAGGTTCCAGGAGTACTTGCAGCGCCAG	974
Db	766	CTGCTGCCCGGACACAAATGTGTGAACAGCAACAGGTTCCAGGAGTACTTGCAGCGCCAG	825
Qy	975	GGCGGTGGCGCGTGTGGGGCAGCGTCACTACCGTCCATGAGAGGGCCATCGCCAC	1034
Db	826	GGCGGTGGCGCGTGTGGGGCAGCGTCACTACCGTCCATGAGAGGGCCATCGCCAC	885
Qy	1035	GCCATGAAGTGGGCCAGGACCAAGTGGAGCTCGCAAGAACAGCTTTTGAGCTCTACGGG	1094
Db	886	GCCATGAAGTGGGCCAGGACCAAGTGGAGCTCGCAAGAACAGCTTTTGAGCTCTACGGG	945
Qy	1095	GCTGACTTCGTTCTGGGAGGGAATTTCAGGCCCTTGGTGTGATTCGAGATCAATTCAGCCCC	1154
Db	946	GCTGACTTCGTTCTGGGAGGGAATTTCAGGCCCTTGGTGTGATTCGAGATCAATTCAGCCCC	1005
Qy	1155	ACCATGACCCGCTCCAGCCCGGTTCAGGCCCGAGCTGTGTGACAGGTCAGGAGGACAC	1214
Db	1006	ACCATGACCCGCTCCAGCCCGGTTCAGGCCCGAGCTGTGTGACAGGTCAGGAGGACAC	1065
Qy	1215	ATCAAGGTGGCCGTGGACCGCAGCTGTGACATCGGCAATCTTCGAGCTCTCTGTGTGAGGAC	1274

Db	1066		ATCAAGGTTGGCCCGTGGACCGCAGCTGTGATCATCGCAACTTTCAGTCTCTGTGG	---	AG	1121
Qy	1275		CCGGTGTGTTGAGCCGCCCCCATTTACGCGGTCGGA	CTCTGCGTGGCGGCGCTCAGTGTG		1334
Db	1122		CCGGTGTGTTGAGCCGCCCCCATTTACGCGGTCGGA	CTCTGCGTGGCGGCGCTCAGTGTG		1181
Qy	1335		AGGAGAGCCAGGAGGAGGTGCTGCCCGTCTGCAAC	CTCAAGGCTTCGCGCTCGCTGTGTTG		1394
Db	1182		AGGAGAGCCAGGAGGAGGTGCTGCCCGTCTGCAAC	CTCAAGGCTTCGCGCTCGCTGTGTTG		1241
Qy	1395		GAGCGGAGCCGCTGAAGGCA	CGGGGCGCTTCGGCATGCCAGACCTGCGCCAGGACCC		1454
Db	1242		GAGCGGAGCCGCTGAAGGCA	CGGGGCGCTTCGGCATGCCAGACCTGCGCCAGGACCC		1301
Qy	1455		CCATCACCACTCTCCAGCGGGACTTTGGGACTTGA	AGGAAAGAGAGGGGCTCCCCCTGGCC		1514
Db	1302		CCATCACCACTCTCCAGCGGGACTTTGGGACTTGA	AGGAAAGAGAGGGGCTCCCCCTGGCC		1361
Qy	1515		TTGCTGGCACCCCTTAAGGGGGGCGAGCGGTGGAG	CGGTGGAGCGCACAGGCCACCCCGCACC		1574
Db	1362		TTGCTGGCACCCCTTAAGGGGGGCGAGCGGTGGAG	CGGTGGAGCGCACAGGCCACCCCGCACC		1421
Qy	1575		AAAGCTGCTGGGAAGGTGAGCTTCCGGGCTTGGCC	CTGTGCCACGTGGAAGTCAAGGCC		1634
Db	1422		AAAGCTGCTGGGAAGGTGAGCTTCCGGGCTTGGCC	CTGTGCCACGTGGAAGTCAAGGCC		1481
Qy	1635		CCAAACACCGTGTCTCCCGTGGCCAGCGCCGCA	AAAGCTGGGATCCAAACCAAGCTAAAT		1694
Db	1482		CCAAACACCGTGTCTCCCGTGGCCAGCGCCGCA	AAAGCTGGGATCCAAACCAAGCTAAAT		1541
Qy	1695		GCGCACCGCTGGAGGCTGTGCTCGCGGGGCTTGA	AGCAGCAGAGGGCGCGCTGCGTCCG		1754
Db	1542		GAGCACCGCTGGAGGCTGTGCTCGGAGGCTTGA	AGCAGCAGAGGGCGCGCTGCGTCCG		1601
Qy	1755		CGCCCCGAGGAAAAAGTTCA	TGACACGCTCAGATTCTGTGACAGAGGATACAGTTGC		1814
Db	1602		CGCCCCGAGGAAAAAGTTCA	TGACACGCTCAGATTCTGTGACAGAGGATACAGTTGC		1661
Qy	1815		AGCCACTCTCCC	1826		
Db	1662		AGCCACTCTCCC	1673		
RESULT 6						
ADJ93359						
ID ADJ93359 standard; cDNA; 726 BP.						
XX	AC	ADJ93359;				
XX	AC					
XX	DT	06-MAY-2004 (first entry)				
XX	XX	Human BGS-42 cDNA sequence SeqID3.				
XX	XX	testis-specific tubulin tyrosine-ligase-like polypeptide;				
KW	KW	BGS-42 polypeptide; cytostatic; respiratory-Gen; gastrointestinal-Gen;				
KW	KW	neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;				
KW	KW	osteopathic; neotropic; antiparkinsonian; antiarthritic; antiasthmatic;				
KW	KW	anti-Hiv; antibacterial; immunosuppressive; antiseborrheic;				
KW	KW	dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;				
KW	KW	tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;				
KW	KW	testicular disorder; testicular cancer; pulmonary disorder; lung cancer;				
KW	KW	gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;				
KW	KW	brain cancer; liver cancer; proliferative condition; testis; lung;				
KW	KW	small intestine; brain; lymph tissue; infertility; Cushing's syndrome;				
KW	KW	emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;				
KW	KW	Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;				
XX	XX	sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.				
OS	OS	Homo sapiens.				
XX	XX					
Key	Key	Location/Qualifiers				
FT	FT	1. .726				
CD5	CD5					


```
/*tag= a
/product= "Human BCS-42 protein"
/partial
/note= "No start or stop codon"
```

WO2004005487-A2.

15-JAN-2004.

09-JUL-2003; 2003WO-US021605.

09-JUL-2002; 2002US-0394725P.

(BRIM) BRISTOL-MYERS SQUIBB CO.

Feder JN, Wu S, Nelson TC;

WPI; 2004-099381/10.

New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, useful for preventing, treating or ameliorating a medical condition, e.g. aberrant cellular proliferation, reproductive disorders or testicular disorders.

Disclosure; SEQ ID NO 3; 343pp; English.

This invention relates to a novel testis-specific tubulin tyrosine-ligase-like polypeptide, designated the BGS-42 polypeptide. The invention may be useful for the development of compounds with a cytostatic, respiratory-gen, gastrointestinal-gen, neuroprotective, endocrine-gen.

-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, antiinflammatory, anabolic, hypertensive, osteopathic, nootropic, antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial, immunosuppressive, antiseborrheic or dermatological activity acting as tyrosine ligase modulators. In addition, the disclosed sequences may be used for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of a partial cDNA which encodes the human BGS-42 protein of the invention.

Sequence 726 BP; 157 A; 227 C; 214 G; 128 T; 0 U; 0 Other;

```
Query Match      39.5%; Score 726; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.6e-145;
Matches 726: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

549 ATTGACGGGCTCGGAACATCTGGATTATAAAGCCCGCGGCCAAGTCCCGGGGCGGAGAC 608
1 ATTGACGGGCTCGGAACATCTGGATTATAAAGCCCGCGGCCAAGTCCCGGGGCGGAGAC 60

609 ATAGTGTGATGGACCGCTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCACCCCTCTT 668
|||||
61 ATAGTGTGATGGACCGCTGTGGAGGAGATCCTGGAGCTGGCAGCTGCAGACCACCCCTCTT 120

669 TCCAGGACAACAAGTGGTGGTCCAGAAGTACATCAGAGACGCCGCTGTCATCTGTGAC 728
|||||
121 TCCAGGACAACAAGTGGTGGTCCAGAAGTACATCAGAGACGCCGCTGTCATCTGTGAC 180

729 ACCAAGTTCGACATCAGACAGTGGTTCCTCGTCACGSACTGGAACCCCTGACCATCTGG 788

Db	181	ACAAAGTTGCAATCAGACAGTGGTTCCTCGTCACGGACTGGAAACCCCTGACCACCTGG	240
Qy	789	TTCTACAAGGAGAGTACTTGGCGTTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC	848
Db	241	TTCTACAAGGAGAGTACTTGGCGTTTCTCAACTCAGCGCTTCTCCCTGGACAAGCTGGAC	300
Qy	849	AGGCGCATCGACTGTGCGCAACAAAGCGCGTCCAGAAGTACCTGAAGAATGATGTGGCGCG	908
Db	301	AGGCGCATCGACTGTGCGCAACAAAGCGCGTCCAGAAGTACCTGAAGAATGATGTGGCGCG	360
Qy	909	AGGCGCGTCTGCGCCGCAACACATGTGACACAGACAGGTTCCAGGAGTACCTGTCAG	968
Db	361	AGGCGCGTCTGCGCCGCAACACATGTGACACAGACAGGTTCCAGGAGTACCTGTCAG	420
Qy	969	CGCCAGGCGCGTGGCGCGTGTGGGCGAGCGTCACTACCCGTCATGAAGAAGGCCATC	1028
Db	421	CGCCAGGCGCGTGGCGCGTGTGGGCGAGCGTCACTACCCGTCATGAAGAAGGCCATC	480
Qy	1029	GCCACGCCATGAAGTGGCGCCAGGACAGCGTGGAGCGCTCGCAAGAACAGCTTTGAGCTC	1088
Db	481	GCCACGCCATGAAGTGGCGCCAGGACAGCGTGGAGCGCTCGCAAGAACAGCTTTGAGCTC	540
Qy	1089	TACGGGCGTCACTTCGTCCTTGGGAGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCC	1148
Db	541	TACGGGCGTCACTTCGTCCTTGGGAGGGACTTCAGGCGCTGGCTGATCGAGATCAATTCC	600
Qy	1149	AGGCGCCACCATGCAACCCGTCACGCGCGTTCAGGCGCGAGCTGTGTGCACAGGTGCAGGAG	1208
Db	601	AGGCGCCACCATGCAACCCGTCACGCGCGTTCAGGCGCGAGCTGTGTGCACAGGTGCAGGAG	660
Qy	1209	GAACCATCAAGTGGCGCGTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG	1268
Db	661	GAACCATCAAGTGGCGCGTGGACCCGAGCTGTGACATCGGCAACTTCGAGCTCCTGTGG	720
Qy	1269	AGGCAG 1274	
Db	721	AGGCAG 726	
RESULT 7			
ADQ17814/c			
ID ADQ17814 standard; DNA; 101270 BP.			
XX	ADQ17814;		
XX	26-AUG-2004 (first entry)		
XX	Human soft tissue sarcoma-upregulated DNA - SEQ ID 631.		
DE	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;		
XX	db.		
OS	Homo sapiens.		
XX	WO2004048938-A2.		
XX	10-JUN-2004.		
XX	26-NOV-2003; 2003WO-US039193.		
XX	26-NOV-2002; 2002US-0429739P.		
XX	(PROT-) PROTEIN DESIGN LABS INC.		
XX	Aziz N, Ginsburg WM, Zlotnik A;		
XX	WPI; 2004-441208/41.		
XX	Early detection of soft tissue sarcoma comprises determining expression		
PT	of a gene in a first soft tissue sample and a normal soft tissue sample		
PT	and comparing the gene expression, also useful in treating soft tissue		
PT	sarcoma.		
XX			

PS Example 2; SEQ ID NO 631; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual

CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level

CC of protein expression in the first soft tissue sample indicates the

CC presence of soft tissue sarcoma. The method of the invention has

CC cytostatic applications and may be useful for detecting soft tissue

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic

CC acid sequences may be useful in diagnostic and screening applications.

CC The current sequence is that of a human soft tissue sarcoma-upregulated

CC DNA of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.

XX

SQ Sequence 101270 BP; 24151 A; 28044 C; 28100 G; 20975 T; 0 U; 0 Other;

Query Match 27.3%; Score 501; DB 12; Length 101270;

Best Local Similarity 100.0%; Pred. No. 6.3e-97;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1272 CAGCCGGTGTTCAGCGCGCCCATTCAGCGGTCCGACCTCTGCTGGGGGCGTCAGT 1331

DB 43637 CAGCCGGTGTTCAGCGCGCCCATTCAGCGGTCCGACCTCTGCTGGGGGCGTCAGT 43578

QY 1332 GTGAGGAGAGCCAGGAGGAGTGTGCTGCCGCTCTGCAACCTCAAGGCTTCGGGCTCGCTG 1391

DB 43577 GTGAGGAGAGCCAGGAGGAGTGTGCTGCCGCTCTGCAACCTCAAGGCTTCGGGCTCGCTG 43518

QY 1392 TTGGAGCGCGAGCGCTGAAGGACACGGGGCCCTCGGCCATGCGACACCTGCCCCAGGGA 1451

DB 43517 TTGGAGCGCGAGCGCTGAAGGACACGGGGCCCTCGGCCATGCGACACCTGCCCCAGGGA 43458

QY 1452 CCCCCATCACCACTCTCCAGCGGACTTGGGACTGAGAGAGAGGGGCTCCCCCTG 1511

DB 43457 CCCCCATCACCACTCTCCAGCGGACTTGGGACTGAGAGAGAGGGGCTCCCCCTG 43398

QY 1512 GCCTTGTGTCACCTTTAAGGGGGGCGAGCGAGCGGTGGAGCCGACAGCCACCCCG 1571

DB 43397 GCCTTGTGTCACCTTTAAGGGGGGCGAGCGAGCGGTGGAGCCGACAGCCACCCCG 43338

QY 1572 ACCAAGAGTGTGGGAAGTGTGGAGTCTCCGGCTGCTGCCCTGTGCTGCCACGTGACAGTCA 1631

DB 43337 ACCAAGAGTGTGGGAAGTGTGGAGTCTCCGGCTGCTGCCCTGTGCTGCCACGTGACAGTCA 43278

QY 1632 GCCCCAAACACCGGTGTCCCGTAGCCAGCCCGCCAAAGCTGGGATCCAAACCAAGCTA 1691

DB 43277 GCCCCAAACACCGGTGTCCCGTAGCCAGCCCGCCAAAGCTGGGATCCAAACCAAGCTA 43218

QY 1692 AATGCGCACCCGCTGAGAGCTGTGCTGCGGGGCTTGAAGACAGACAGAGGGCGGCTGCGCT 1751

DB 43217 AATGCGCACCCGCTGAGAGCTGTGCTGCGGGGCTTGAAGACAGACAGAGGGCGGCTGCGCT 43158

QY 1752 CCGCCCGCCGGAGGAAAAGT 1772

DB 43157 CCGCCCGCCGGAGGAAAAGT 43137

RESULT 8

AAS99894

ID AAS99894 standard; cDNA; 2380 BP.

XX

AC AAS99894;

XX

DT 12-MAR-2002 (first entry)

XX

DE Polynucleotide encoding human cytoskeleton-associated protein (CYSKP) #5.

XX

KW Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;

KW cell proliferative disorder; inflammatory disorder; prion disease;

KW vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;

KW neurological disorder; cell motility disorder; reproductive disorder;

KW spinal cord disease; central nervous system disorder; mental disorder;

KW gene therapy; cancer.

XX

OS Homo sapiens.

XX WO200185942-A2.

PN

XX 15-NOV-2001.

XX

XX 03-MAY-2001; 2001WO-US014355.

XX

XX 05-MAY-2000; 2000US-0201960P.

PR 08-MAY-2000; 2000US-0202729P.

PR 05-JUN-2000; 2000US-0209705P.

PR 07-JUN-2000; 2000US-0210149P.

PR 21-JUN-2000; 2000US-0213215P.

XX

XX (INCY-) INCYTE GENOMICS INC.

PA

XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;

PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;

PI Policky JL;

XX WPI; 2002-062248/08.

DR P-PSDB; AAU74334.

XX

XX New cytoskeleton-associated proteins and polynucleotides, useful for

PT diagnosing, preventing and treating cell proliferative, autoimmune,

PT inflammatory, neurological, cell motility, reproductive and muscle

PT disorders.

XX

PS Claim 5; Page 171; 194pp; English.

XX

CC The invention relates to human cytoskeleton-associated polypeptides

CC (CYSKP) and their associated polynucleotide sequences. The sequences are

CC useful in the treatment of disorders associated with overexpression or

CC underexpression of CYSKP in a patient. The disorders include cell

CC proliferative disorders (such as cancer, actinic keratosis,

CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),

CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,

CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus

CC and anaemia), vesicle trafficking disorders (such as

CC hypercholesterolaemia, diabetes insipidus, Grave's disease and goitre),

CC gastrointestinal disorders, prion diseases, neurological disorders (such

CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,

CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis

CC and other motor neuron disorders), cell motility disorders, reproductive

CC disorders (such as endometriosis and polycystic ovary syndrome), muscle

CC disorders (such as myocarditis, migraine, hypertension, hypoglycaemia,

CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord

CC diseases, central nervous system disorders (such as Down syndrome and

CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).

CC Sequences AAS99890-AAS99923 represent cDNA molecules encoding human CYSKP

CC of the invention

XX

SQ Sequence 2380 BP; 580 A; 685 C; 670 G; 445 T; 0 U; 0 Other;

Query Match 23.5%; Score 432.8; DB 6; Length 2380;

Best Local Similarity 59.1%; Pred. No. 1.1e-82;

Matches 802; Conservative 0; Mismatches 542; Indels 14; Gaps 3;

QY 260 CAGCAGGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCCTCCGGGGCAGCTTG 318

DB 361 CAGGAGACAAGCAGAGCCCAAGAACAGGAGAAAAACCCAGTGTGGTGTCCCCAGAGTTTG 420

QY 319 TGGACATCGCTGCAAGGTGTCCAGGCTACTCTGGGCGAGCTGGAGCATGAGGACATCG 378

DB 421 TGGATGAAGCTCTGTGTGCTGCGGAGGTACTCTTAGCAACTTGGCCCAATGAGCATCG 480

QY 379 ACACGCTCAGCAGATGCTGCGGAGGAGCTTCACTGAGGCCGAGTGGGAGGACCTGACCCAGC 438

DB 481 ACAGAGCCTGGAGGCGCCGCTGTACTCACCCCGAGGGCTGTGTCCTCTTCTCCACG 540

QY 439 AGTACTACTCCCTGTTTCATGGCGATGCTTTTCATCTCCAAATTCAGAAATTTACTTTTCGC 498
Db |||||
QY 541 GCTACTACCAAGTGTCTCAAGAGGGGAGAGACTCAGGCACCTCGACACTCAGTCCAGC 600
Db |||||
QY 499 AGTGCCAGGCTCTGCTGAATAGAAATCATGCTGTGTGAACCCCTCAGACGGACATTTGACGGGC 558
Db |||||
QY 601 GCTGTGAGGACATCTCGCAGCAGCTGCAGGCGGTGGTACCCACATAGACATGGAAGGGG 660
Db |||||
QY 559 TCCGGACATCTGATTTATAAGCCCGGCGCAAGTCCCGGGGCGGAGACATAGTGTGCA 618
Db |||||
QY 661 ATCCCAACATCTGGATCGTGAAGCCAGGAGCAAGTCCCGTGGACGAGCATCATGTGCA 720
Db |||||
QY 619 TGGACCGTGTGGAGCAGATCTCGAGCTGGCAGCTGCAGACCCCTCTTTCCAGGGACA 678
Db |||||
QY 721 TGGACCACTGGAGAGATCTGAGCTGTGAACGGCAACCCGCTGTATGAGAGCG 780
Db |||||
QY 679 ACAAGTGGGTGTTCAGAAATACATCGAGAGCGCGCTGCTCATCTGTGACACCAAGTTGC 738
Db |||||
QY 781 GCAAGTGGGTGTTCAGAAATATATTGAGCGGCCCTCTCATCTTTGGCACCAAGTTTG 840
Db |||||
QY 739 ACATCAGACAGTGGTTCCTGTACGAGCTGGAACCCCTTGACCATCTGTTCTACAAAG 798
Db |||||
QY 841 ACCTCAGACAGTGGTTCCTGTAACTGAACTGGAACCCCACTTACCGTGTGTTCTACCGCG 900
Db |||||
QY 799 AGAGTTACTTTCGGTCTCAACTCAGCGCTTCTCCTGGCAAGCTGCAGCGCCATCC 858
Db |||||
QY 901 ACAGCTATATCCGCTTTTCCAGCGAGCCCTTCTCCTTGAAGACCTGGAACACTCAGTGC 960
Db |||||
QY 859 ACCTGTGCAACAAACCGCGTCCAGAAATACCTGGAAGAAATGATGTGGGCGCGAGCCCTCTGC 918
Db |||||
QY 961 ACCTGTGCAACAACTCCATCCAGAAACACCTGGAGAACTCATGCCATCGGCATCCACTGC 1020
Db |||||
QY 919 TGCCCGCACACATGTGACACAGCAACAGGTTTCCAGAGTACCTGCGAGCGCCAGGGCC 978
Db |||||
QY 1021 TTCCCGCAGACAAACATGTGTCTAGCCAGAGTTTCCAGGGCCCACTGCGAGAGATGGGTG 1080
Db |||||
QY 979 GTGGCGCGGTGGGGCAGGTCTATCTACCGTCCATGAAGAGGCGCATCGCCACGCCA 1038
Db |||||
QY 1081 CCCCAATGCTTGTGTCACCATCATCTGCTGCTGATGAAGGATGCTGTATCCAGCAC 1140
Db |||||
QY 1039 TGAAGTGGCGCCAGGACACAGTGGAGCTCGCAAGAACAGCTTTGAGCTCTAGGGGCTG 1098
Db |||||
QY 1141 TTGAGACCTCCAGGACACCGTGCAGTGTGGAAGGCCAGCTTTGAGCTATATGGCGCTG 1200
Db |||||
QY 1099 ACTTGTCTTGGAGGAGATTTAGCCGCTTGGCTGTGATCGAGATCAATTCAGGCCCAACA 1158
Db |||||
QY 1201 ACTTGTGTTCGGGAGGACTTCAGCCCTGGCTGATTGAGATCAAGCCAGGCCCAACA 1260
Db |||||
QY 1159 TGCACCGTCCAGCGGTCAAGCCCGCAGCTGTGTGCACAGGTGCAGGAGGACCATCA 1218
Db |||||
QY 1261 TGGCACCTCCACAGCAGTCACTGCCCGGCTGTGTGCTGGCGTGCAGCTGACACCCCTGC 1320
Db |||||
QY 1219 AGGTGGCC-----GTGACCGCAGCTGTGACATCGGCAACTTCGAGCTCTCTGT 1266
Db |||||
QY 1321 GCGTGGTCAATTGACCGGATGCTGGACCGCACTGTGACACAGAGCCTTTGAGCTCATCT 1380
Db |||||
QY 1267 GGAGGACGCGGTGTGAGCCGCCCATTTACGCGGTCGACCTCTGCGTGGCGGGG 1326
Db |||||
QY 1381 ATAAGCAGCCTGCTGTGAGGTGCTCAATATGTGGGCATCCGGCTCTCTGTTAGAGGCT 1440
Db |||||
QY 1327 TCAGTGTGAGGAGCCAGGAGCAGTGTGCTGCCGTCTCAACCTCAAGGCTCTCGCCT 1386
Db |||||
QY 1441 TCACCATCAAGAGCCCATCGCATGTGTATCGGCGGATGGGGGTCCGCCACGAGTCC 1500
Db |||||
QY 1387 CGCTGTTGGACGCGCAGCCGCTGAAGCAGCGGGGCCCTCGGCCATGCC-AGACCCCTGCC 1445
Db |||||
QY 1501 CTCTGTGACCCAGCGAGGCTCTGGGGAAGGCAAGACTCGGGATCCCTACCCACAGGT 1560
Db |||||
QY 1446 CAGGAGCCCCCATCAACAGCTCTTCCAGCGGACTTGGGACTGTAAGGAAGAGAGGGGCTC 1505
Db |||||
QY 1561 CAGCTTCTAGGAAGGCACTGGGGCCAGGAGCCTTGGGGCAGTGTGAGAGCCAGTCTCCA 1620
Db |||||
QY 1506 CCCCTGCGCTTGTGGCACCCCTTAAGGGGGGCGAGCGGAGAGCGGTGGAGCGGCACAGCCC 1565
Db |||||

Db 1621 CTGCCACCACTTCAGCCCGGAAGGGGAAGGCAAGGCGAAGGGCCACACGCC 1680
QY 1566 ACCCGCACCAAGCTCTCTGGGAAGGTGGAGCTCCCGGC 1603
Db 1681 TGGTCTGCCCAATCTCTGGAGTGGGATGCCCCAGC 1718
RESULT 9
AAH16735
ID AAH16735 standard; cDNA; 2326 BP.
XX
AC AAH16735;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15920.
XX
KW Human; primer; detection; diagnosis; antiseize therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000BP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 8; SEQ ID NO 15920; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 2326 BP; 524 A; 701 C; 642 G; 459 T; 0 U; 0 Other;

Query Match		23.5%; Score 431.6; DB 4; Length 2326;
Best Local Similarity		63.5%; Pred. No. 2e-82;
Matches		698; Conservative 0; Mismatches 389; Indels 13; Gaps 2;
QY	260	CAGCAGCAAGATGCTGAAAATGCTGAGGCAAAAGCTCAG-GGGCCTCCGGGCGCAGCTTG 318
DB	620	CAGAGACAAGCAGCCCAAGAAACAGAGAAAACCCAGGTGGTGTCCTCCAGAGATTG 679
QY	319	TGGACATCGCGTCAAGGTGTGCGAGGCTTACCTGGGGCAGCTGGAGCATGAGGACATCG 378
DB	680	TGGATGAAGCTCTGTGCTGCGAGGAGTACCTTAGCAACTTGGCCCATGACATCG 739
QY	379	ACAGGTCAGCAGATCGCGTGGAGGACCTTACTAGGCGCGAGTGGAGGACCTGACCCAGC 438
DB	740	ACAAGGACCTTGGAGGCGCCGCTGTACTACCCCGGAGGCGTGTCTCTTCTCCAGC 799
QY	439	AGTACTACTCCCTCGCTTCATGGGATGCTTTTCTATCTCCRAATTCAGAAATTACTTTTCG 498
DB	800	GCTTACTACCAAGTGGTCCAGAAAGGGCGAGAACTCAGGCACTCTGACACTCAGGTCAGC 859
QY	499	AGTGCCAGGCTCTGCTGAATAGAAATCACGCTCTGTGAACCCCTCAGACGGCAATTGACGGGC 558
DB	860	GCTGTGAGGACATCTCTGCAGCAGCTGCAGCGCTGTGTACCCAGATAGACATGGAAGGG 919
QY	559	TCCGGAACATCTGGATTATAAGCCCGCGGCCCAAGTCCCGGGGCCGAGACATAGTGTGCA 618
DB	920	ATCGCAACATCTGGATCGTAGGCCAGGAGCCAAAGTCCCGCGGACGAGGCATCATGTGCA 979
QY	619	TGGACCGTGTGGAGGAGATCTCGAGCTGGCAGCTGCAGACCACTCTTTCAGGGACA 678
DB	980	TGGACCACTTGGAGGAGATGCTGAAGCTGTGAACGGCAACCCCGTGGTGATGAAGGAGC 1039
QY	679	ACAAGTGGGTGTCCAGAAGTACATCGAGACGCGCTGTCTCATCTGTGACACCAAGTTTCG 738
DB	1040	GCAAGTGGGTGTGCAAGATATATTGAGCGGCCCTCTCTCATCTTTGGCACCAAGTTTG 1099
QY	739	ACATCAGACAGTGGTTCCTCGTCACGAATGGAACCCCTGACCATCTGTTTCTACAGG 798
DB	1100	ACCTCAGACAGTGGTTCCTGTTAACTGACTGGAACCCACTTACCGTGTGTTCTACCGCG 1159
QY	799	AGAGTTACTTGGCGTCTCAACTCAGCGTTCTCCCTGGACAGCTGGACGGCCATCC 858
DB	1160	ACAGCTATATCCCGCTTTTCCAGCAGCCCTTCTCCCTGAAGAACCTTGGACAACTCAGTGC 1219
QY	859	ACCTGTGCAACACGCGCTCCAGAAGTACCTGAAGAATGATGTGGCGCAGCCCCCTCTGC 918
DB	1220	ACCTGTGCAACACTCCATCCAGAACACTCTGAGAACTCATGCCATCGGCATCCACTGC 1279
QY	919	TGCCCGCACACAATGTGGACCAAGCACAGGTTCCAGAGTACCTGACGGCCAGGGCC 978
DB	1280	TTCCGCGCAGACAATGTGGTCTAGCCAGAGGTTCCAGGCCCCACCTGCGAGGATGGGTG 1339
QY	979	GTGGCGCGCTGTGGGCGAGGTCTATCCCGTCCATGAGAGGCCATCGGCCACGCCA 1038
DB	1340	CCCCAAATGTTGTTCCACCATCATCTGCTGGCATGAAGGATGTGTGATCCAGCAC 1399
QY	1039	TGAAGTGGGCCAGGACCAAGTGGAGCTTCGCAAGAACAGCTTTGAGCTCTACGGGGCTG 1098
DB	1400	TTCAGACCTCCAGGACACGCTGACGTGTCGGAAGCCAGCTTTGAGCTCTATGGCGCTG 1459
QY	1099	ACTTGTCTCTTGGAGGAGACTTCAGGCGCTTGGCTGTGATTCAGATCAATTCAGCCCCACCA 1158
DB	1460	ACTTGTGTTCGGGGAGGACTTCAGGCGCTTGGCTGTGATTCAGATCAACGCGCAGCCCCACGA 1519
QY	1159	TGCACCGCTCCAGCGGTTCAGGCGCCAGCTGTGTCACAGGTGCGAGGAGCACCATCA 1218
DB	1520	TGGCACTCTCCAGCAGTCACTGCGCCGGCTCTGTGCTGGCGTGAAGCTGACACCTGTC 1579
QY	1219	AGGTGGCC-----GTGGACCGCAGCTGTGACATTCGGCAACTTCGAGCTCCTGT 1266
DB	1580	GCGTGTCTATTGACCGGATGCTGACCGCAACTGTGACACAGGAGCTTTGAGCTCATCT 1639
QY	1267	GGAGGCGAGCGGTGTTGAGCGCCGCCCATTTACGCGGCTCCGACCTCTGCGTGGCGGGCG 1326

DB	1640	ATAAGCAGCTCTGCTGGAGTGCTTCAATATGTGGGCATCGCGCTCCTGGTAGAGGCT 1699
QY	1327	TCAGTGTGTGAGGAGCCAGG 1346
DB	1700	TCACCATCAAGAGCCCATG 1719
RESULT 10		
ACH91699/c		
ID	ACH91699	standard; DNA; 490 BP.
XX	ACH91699;	
XX	29-JUL-2004	(first entry)
XX	Human genome derived single exon probe #24894.	
XX	Human; probe; ss; gene expression; single exon probe; microarray;	
KW	alternative splicing event; genomic alteration.	
XX	Homo sapiens.	
XX	US2003194704-A1.	
XX	16-OCT-2003.	
XX	03-APR-2002; 2002US-00029386.	
XX	03-APR-2002; 2002US-00029386.	
XX	(PENN/) PENN S G.	
PA	(RANK/) RANK D R.	
PA	(HANZ/) HANZEL D K.	
XX	Penn SG, Rank DR, Hanzel DK;	
XX	WPI; 2004-119264/12.	
XX	New human genome-derived single exon nucleic acid probes useful for human	
PT	gene expression analysis, for identifying or characterizing alternative	
PT	splicing events, for assessing genomic alterations or as tools for	
PT	surveying tissues.	
XX	Claim 1; SEQ ID NO 24894; 80pp; English.	
XX	The invention relates to a nucleic acid probe for measuring human gene	
CC	expression, comprising any of the 27,400 fully defined nucleotide	
CC	sequences in the specification, or their complements or fragments, and	
CC	encoding at least 8 amino acids of any of the 688 amino acid sequences	
CC	fully defined in the specification. The probe is a single exon probe that	
CC	hybridizes under high stringency conditions to a nucleic acid molecule	
CC	expressed in human cells or tissues. Also included are a spatially-	
CC	addressable set of single exon nucleic acid probes for measuring human	
CC	gene expression (comprising a plurality of single exon nucleic acid	
CC	probes cited above, where each of the plurality of probes is separately	
CC	and addressably isolatable or amplifiable from the plurality), a single	
CC	exon microarray for measuring human gene expression, a method of	
CC	measuring human gene expression, a vector comprising the single exon	
CC	probe cited above, an ORF-encoded peptide comprising at least 8	
CC	contiguous amino acids of any of the above-mentioned amino acid	
CC	sequences (optionally with conservative amino acid substitutions), an	
CC	isolated antibody that binds specifically to a peptide cited above,	
CC	methods of selling and/or licensing single exon probes or microarrays to	
CC	a customer desiring to measure gene expression, a method of providing	
CC	human gene expression data by subscription, and a computer-readable	
CC	storage medium which contains a database having a plurality of records	
CC	(each record including data on the expression of a single exon probe	
CC	cited above. The probe, methods and apparatus are useful in gene	
CC	expression analysis. The probes may be used as tools for surveying	
CC	tissues to detect the presence of expressed messages that contain their	
CC	specific exon, or in constructing genome-derived single exon microarrays.	
CC	In addition, the probes are used in identifying and characterizing	

CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ Sequence 490 BP; 77 A; 150 C; 168 G; 95 T; 0 U; 0 Other;

Query Match 23.2%; Score 427; DB 12; Length 490;
Best Local Similarity 100.0%; Pred. No. 1.3e-81;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 848 CAGGGCCATCCACTGTGCAACAAACGCGTCCAGAGTACTGAGAAATGATGTGGCCG 907
DB 427 CAGGGCCATCCACTGTGCAACAAACGCGTCCAGAGTACTGAGAAATGATGTGGCCG 368
QY 908 CAGCCCCCTGCTGCCGCGCACACACATGTGGACCGACACAGTTCCTCAGAGTACCTGCA 967
DB 367 CAGCCCCCTGCTGCCGCGCACACACATGTGGACCGACACAGTTCCTCAGAGTACCTGCA 308
QY 968 CGCCACAGGGCGGTGGCGCCGTGTGGGCGAGCGTCATCTACCCGTCCATGAAGAAGGCCAT 1027
DB 307 CGCCACAGGGCGGTGGCGCCGTGTGGGCGAGCGTCATCTACCCGTCCATGAAGAAGGCCAT 248
QY 1028 CGCCACAGCGCATGAAGTGTGCCAGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCT 1087
DB 247 CGCCACAGCGCATGAAGTGTGCCAGGACCAACGCTGGAGCCTCGCAAGAACAGCTTTGAGCT 188
QY 1088 CTACGGGGCTGACTTCGTCTCGGAGGACTTCAGGCGCTGCTGATCAGATCAATTC 1147
DB 187 CTACGGGGCTGACTTCGTCTCGGAGGACTTCAGGCGCTGCTGATCAGATCAATTC 128
QY 1148 CAGCCCCACCATGACACCCGTCACCGCGTCCAGGCGCCAGCTGTGTGCACAGGTGACGA 1207
DB 127 CAGCCCCACCATGACACCCGTCACCGCGTCCAGGCGCCAGCTGTGTGCACAGGTGACGA 68
QY 1208 GGAACACATCAAGGTGGCGCTGTGACCGCAGCTGTGACATCGGCAACTTCGAGTCTCTGTG 1267
DB 67 GGAACACATCAAGGTGGCGCTGTGACCGCAGCTGTGACATCGGCAACTTCGAGTCTCTGTG 8
QY 1268 GAGGCAG 1274
DB 7 GAGGCAG 1

RESULT 11

AD573168
ID AD573168 standard; cDNA; 2538 BP.
XX
AC ADS73168;
XX
AT 18-NOV-2004 (first entry)
XX
DE Human kidney tumour specific cDNA, SEQ ID 1765.
XX
KW Human; ss; kidney tumour; kidney cancer; cytostatic; gene therapy;
KW T-cell; immune response.
XX
OS Homo sapiens.
XX
FN US2003109434-A1.
XX
PD 12-JUN-2003.
XX
PF 19-MAR-2002; 2002US-00102524.
XX
PR 19-MAR-2001; 2001US-0277245P.
PR 21-DEC-2001; 2001US-0343340P.
XX
PA (CORI-) CORIXA CORP.

XX
PI Algate PA, Mannion J, Gaiger A, Gordon B, Harlocker SL;
XX WPI; 2002-759855/82.
XX
PT New isolated polynucleotides and polypeptides, useful for detecting the
PT presence of, and treating cancer, particularly kidney cancer by
PT stimulating T-cells specific for a tumor protein, and stimulating immune
PT response in a patient.
XX
XX Claim 1; SEQ ID NO 1765; 78pp; English.
XX
XX The invention relates to a new isolated polynucleotide (a Human kidney
XX tumour specific cDNA) comprising any one of the 1855 sequences identified
XX in the specification (or their complements, degenerate variants,
XX sequences consisting of at least 20 contiguous residues them, sequences
XX that hybridise to them under highly stringent conditions or sequences
XX having at least 75 or 90% sequence identity to the 1855 sequences. Also
XX included are detecting/determining the presence of cancer in a patient,
XX stimulating an immune response in a patient; treating kidney cancer in a
XX patient, an isolated polypeptide encoded by one of the 1855 sequences, an
XX expression vector comprising the polynucleotide operably linked to an
XX expression control sequence, a host cell transformed/transfected with the
XX vector, an isolated antibody (or its antigen-binding fragment) that
XX specifically binds to the protein, a fusion protein comprising at least
XX one the proteins, stimulating and/or expanding T-cells specific for a
XX tumour protein, an isolated T-cell population comprising the T-cells, a
XX composition comprising a first component (such as a carrier or
XX immunostimulant) and a second component (comprising one of the
XX polynucleotides, the polypeptides, an antibody, T-cell or an antigen-
XX presenting cell that expresses the polynucleotide) and a diagnostic kit
XX comprising at least one of the oligonucleotides, or at least one antibody
XX and a detection reagent comprising a reporter group. The polynucleotides,
XX polypeptides, antibodies and antigen-presenting cells are useful for
XX detecting the presence of, and treating cancer, particularly kidney
XX cancer by stimulating and/or expanding T-cells specific for a tumour
XX protein, and stimulating immune response in a patient. The present
XX sequence is one of the Human kidney tumour specific cDNAs. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from USPTO
XX at seqdata.uspto.gov/sequence.html?DocID=20030109434.
XX
SQ Sequence 2538 BP; 555 A; 717 C; 734 G; 532 T; 0 U; 0 Other;
Query Match 22.9%; Score 421.4; DB 7; Length 2538;
Best Local Similarity 63.8%; Pred. No. 3e-80;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 232 GGGAGGAGCGCGGAGCAGGACCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAA 291
DB 754 GGGATCGTACAGGCGCCCTCCCTCTATGTGAGGAGACAGCAGCCCAAGAAACAGGAGAA 813
QY 292 AGCTCAG-GGGGCTCCCGGGGCGAGCTTGTGGACATCGCGTGCAGAGTGTGCCAGGCGCTAC 350
DB 814 AACCCAGTGTGGTGTCCCGAGAGTTTGTGATGAAGCTCTGTGCGTGCAGAGAGTAC 873
QY 351 CTGGGGCAGCTGAGCATGAGGACATCGACACGTCAGCAGATGCCCTGGAGGACCTCACT 410
DB 874 CTAGCAACTTGGCCCAATGAGACATCGACAGGACCTGGAGGCGCCGCTGTACCTCACC 933
QY 411 GAGGCGAGTGGGAGGACCTGACCCAGCAGTACTCTCCCTCGTTCATGCGGATGCTTTC 470
DB 934 CCGAGGGCTGTGTCCTCTTCTCCAGCGCTACTACAAAGTGTGCCAAGGGGCGAGAA 993
QY 471 ATCTCCAAATTCAAGAAATTTACTTTTCGAGTGCAGGCTCTGCTGAATAGAATCAAGTCT 530
DB 994 CTCAGGACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTCTGAGGAGCTGCAGGCC 1053
QY 531 GTGAACCTTCAGACGACATTTGACGGGCTCCGGAACATCTGGATTATAAAGCCCGCGGCC 590
DB 1054 GTGGTACCCACAGATAGACATGGAAGGGGATCGCAACATCTGGATCTGTAAGCAGGAGCC 1113
QY 591 AAGTCCCGGGCGGAGACATAGTGTGCATGGACCGTGTGGAGGAGATCTCTGGAGCTGGCA 650

Db 1114 AAGTCCCGGAGAGGAGCATCATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGTG 1173
Qy 651 GCTGCAGACCACTCTTTCCAGGGCAACAAGTGGTGGTCCAGAAATACATCGAGAG 710
Db 1174 AAGCGAACCCTGGTGTATGAAGGAGCGCAAGTGGTGGTGCAGAAATATATTGAGCG 1233
Qy 711 CGCTGCTCATCTGTGACACCAAGTTGCAATCAGACAGTGGTTCTCGTCAGGACTGG 770
Db 1234 CCCTCTCATCTTTGGCAACCAAGTTGACCTCAGACAGTGGTTCTCGTAACTGACTGG 1293
Qy 771 AACCCCTGACCATCTGGTTCTCAAGAGAGATTACTTGGGGTTCTCAACTCAGCGCTTC 830
Db 1294 AACCCACTTACCGTGGTGTCTACCGGACAGCATATCGCTTTTCCACGCGAGCCCTTC 1353
Qy 831 TCCTCGACAAGCTGCAGAGCGCATCCACCTGTGCAACACGCGCTCCAGAGTACCTG 890
Db 1354 TCCTGAAGACCTGACAACTCAGTGCACCTGTGCAACATCCATCCAGAAACCTG 1413
Qy 891 AAGAATGATGTGGGCGCGAGCCCTGCTGCCGACACAAATGTGGACCAACAGG 950
Db 1414 GAGAACTCATGCCATCGGCATCCACTGCTTCGCCAGACAAATGTGGTCTAGCCAGAG 1473
Qy 951 TTCCAGGAGTACTGCAGCGCCAGGCGCTGGCGCGTGTGGGAGAGCTCATCTACCGG 1010
Db 1474 TTCCAGGCGCACCTGCAGGAGATGGGTGGCCCAAAATGCTTGGTCCACCATCATCTG 1533
Qy 1011 TCATCAAGAAGGCCATCGCCAGCCATGAGGTGGCGCCAGGACCACTGGAGCTCGC 1070
Db 1534 GGCAATGAAGATGCTGTGATCCAGCACCTTCAGACCTCCAGAGACACCGTGCAGTGC 1593
Qy 1071 AAGAACAGCTTTGAGCTCTACGGGGTGTACTTCTGCTTGGAGGAGACTTCAGGCCCTG 1130
Db 1594 AAGGCCAGCTTTGAGCTCTATGGCGTGTACTTGTGTGGGAGAGACTTCCAGCCCTG 1653
Qy 1131 CTGATCGAGATCAATTCCAGCCCCACATGACCCCTGCACGCGCGGTCTACGGCCAGCTG 1190
Db 1654 CTGATTGAGATCAACGCCAGCCCCAGGATGGACCTCCACAGCAGTCACTGCGCGCTC 1713
Qy 1191 TGTGCACAGTGCAGGAGACACCATCAAGTGGCC-----GTGGACCGCAGC 1238
Db 1714 TGTGCTGGCGTGCAGCTGACACCTTCGCGTGGTCAATTGACCGGATGCTGGACCGCAAC 1773
Qy 1239 TGTGACATCGGCAACTTTCAGAGCTTCCTGTGGAGGAGCGCCGT 1279
Db 1774 TGTGACAGAGGACCTTTGAGCTCATCTATAGCAGCCCGT 1814

RESULT 12
ID ACN37881
XX ACN37881 standard; cDNA; 2553 BP.
XX ACN37881;
XX
XX
XX 18-NOV-2004 (first entry)
XX
XX Tumour-associated antigenic target (TAT) cDNA DNA324273, SEQ ID NO:1057.
XX
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic; gene; ss.
XX Homo sapiens.
XX
XX WO2004030615-A2.
XX
XX 15-APR-2004.
XX
XX 29-SEP-2003; 2003WO-US028547.

XX 02-OCT-2002; 2002US-0414971P.
XX (GETH) GENENTECH INC.
XX Wu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
XX P-PSDB; ABM80420.
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX Claim 1; SEQ ID NO 1057; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT nucleic acid of the invention
XX
XX Sequence 2553 BP; 575 A; 716 C; 733 G; 529 T; 0 U; 0 Other;
XX
XX Query Match 22.9%; Score 421.4; DB 13; Length 2553;
XX Best Local Similarity 63.8%; Pred. No. 3e-80;
XX Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
XX
Qy 232 GGGAGAGGCGGGAGCAGCGACCTGAGCAGCAGGCAAGATGCTGAAATGCTGAGGCAA 291
Db 754 GGGATCGTGACAGGCCCTCCCTATGTGCGAGAGACAGCAGCCCAAGAAACAGGAGAA 813
Qy 292 AGCTCAG-GGGCCTCCCGGGCAGCTTGTGGACATCGCGTGCAAGTGTGCGAGGCTTAC 350
Db 814 AACCCAGTGTGGTGTGCCAGAGTTTGTGGATGAAGCTCTGTGTCGTGCGAGGAGTAC 873
Qy 351 CTGGGGCAGCTGGAGCATGAGGACATCGACATCGCAGCAGATGCCGTGGAGGACCTCACT 410
Db 874 CTTAGCAACTTGGGCCCAATGACATCGACAGGACCTGGAGGCCGCCCTGTACTCTCAC 933
Qy 411 GAGGCGAGTGGAGGAGCCTGACCAGCAGTACTACTCCCTCGTTTCATGCCGATGCTTTC 470
Db 934 CCGAGAGGCTGTGCTCCTCTTCTCCAGCGCTACTACCAAGTGTGCCAAGAGGGCAGAA 993
Qy 471 APTCCMAATTCAGAAATTAATCTTTTTCGAGTGCAGGCTCTCTGTAATAGAATCAGCTCT 530
Db 994 CTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTGCGAGCAGCTGCAGGCC 1053
Qy 531 GTGAACCTTCAGACGACATTCACGGGCTCCGGAACATCTGGATTTATAAGAGCCCGGCC 590
Db 1054 GTGGTACCCAGATAGACATGGAAGGGGATCCCAACATCTGGATCGTGAAGCAGGAGCC 1113
Qy 591 AAGTCCCGGGCGGAGACATAGTGTGCAATGGACCGCTGTGGAGGAGATCCTGGAGCTGGCA 650
Db 1114 AAGTCCCGGGAGCAGGGCATCATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGTG 1173
Qy 651 GCTGCAGACCACTCTTTTCAGGAGCAACAAAGTGGGTGTCAGAGAGTACATCGAGACG 710

Db 1174 AACGGCAACCCGTTGGTGTATGAGGACGGCAGTGGGTGTGCGAAGTATATTTAGCGG 1233
Qy 711 CCGCTGCTCATCTGTGTGACACAAAGTTCGACATCAGACAGTGGTTCCTCGTCAAGGACTGG 770
Db 1234 CCGCTGCTCATCTTTGGCACCAGTGTGACCTCAGACAGTGGTTCCTGGTAACTGACTGG 1293
Qy 771 AACCCCTGACCATCTGTTGTTCTACAGGAGAGTACTTGGGTTCTCAACTCAGCGCTTC 830
Db 1294 AACCCACTTACCGTGTGGTGTACCGGACAGCTATATCCGCTTTTCCAGCGACCGCTTC 1353
Qy 831 TCCCTGACAAAGTGTGACAGCGGCATCCACTGTGCAACAAACCGCTCCAGAAAGTACTCG 890
Db 1354 TCCCTGAAGAACCTGGACAACTCAGTGCACCTGTGCAAACTCCATCCAGAGACCTG 1413
Qy 891 AAGATGATGTGGCCGACGCCCTGCTGCGCGCACACAACTGTGGACCAACCAAGG 950
Db 1414 GAGAACTCATCGCCATCGGCATCCACTGCTTCCGCGCAGACAACTGTGTTAGCCAGAGG 1473
Qy 951 TCCAGAGTACCTGACAGCGCCAGGGCGGTGGCGCGCTGTGGGCGACGCTCATCTACCG 1010
Db 1474 TCCAGGCGCACCTGGCAGGAGATGGGTGGCCCCAAATGCTTGGTCCACCATCATCTGTCCT 1533
Qy 1011 TCCATGAAGAAGGCCATCGCCCGCCATGAAGTGGCCCGCAGGACCAACGCTGGAGCTTCGC 1070
Db 1534 GGCATGAGGATGCTGTGATCCAGCACTTCAGACCTCCAGACACCGTGCAGTGTGG 1593
Qy 1071 AAGAACAGCTTTGAGCTCTACGGGGTGAATTCGTCTTGGGAGGACTTCAGGCCCTGG 1130
Db 1594 AAGGCCAGCTTTGAGCTCTATGGCGTGAATTCGTGTTCGGGAGGACTTCCAGCCCTGG 1653
Qy 1131 CTGATGAGATCAATTCACGCCCCACCATGACCGCTCCAGCGCGGTTCAGGCCCGCTG 1190
Db 1654 CTGATGAGATCAACGCCAGCCCCAGATGGCACCTCCACAGCAGTCACTGCGCCGCTC 1713
Qy 1191 TGTGACAGTGTGAGGAGACACATCAAGTGTGCGC-----GTGACCGCAGC 1238
Db 1714 TGTGCTGGGTGACAGCTGACCTCGCGTGTGTATGACCGGATGCTGGACCGCAAC 1773
Qy 1239 TGTGACATCGGCAACTTCGAGCTCTCTGTGGAGCAGCGCGT 1279
Db 1774 TGTGACACAGGACCTTTGAGCTCATCTATAGCAGCGCGCT 1814

RESULT 13

AAC77214
ID AAC77214 standard; cDNA; 2881 BP.
XX
AC AAC77214;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2769 polynucleotide sequence SEQ ID NO:5537.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW immunosuppressant; antidiabetic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; SS.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX

PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimketa RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB43005.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 5; Page 4718-4720; 5507pp; English.
XX
CC AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 2881 BP; 640 A; 812 C; 840 G; 589 T; 0 U; 0 Other;

Query Match 22.9%; Score 421.4; DB 3; Length 2881;
Best Local Similarity 63.8%; Pred. No. 3.1e-80;
Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;
QY 232 GGGAGAGGCGGGAGCAGGACCTGAGCAGCAGGCAAGTGTGAAATGCTGAGGCAA 291
DB 754 GGGATCGTGACAGGCCCTCCCTATGTGAGGAGACAGCAGCCCAAGAACAGGAGAAA 813
QY 292 AGCTCAG-AGGGCTCCCGGGCAGCTTGTGGACATCGCTGCAAGGTGTGCCAGGCTAC 350
DB 814 AACCCAGTGTGGTGTCCCGAGAGTTGTGGATGAAGTCTGTGTGCGCGAGGAGTAC 873
QY 351 CTGGGGCAGCTGAGCATGAGGACATCGACACATCGACAGATGCCGTGGAGGACCTCACT 410
DB 874 CTTAGCAACTTGGCCCATGACATCGACAGGACCTGGAGGCCCGCTGTACTCACC 933
QY 411 GAGGCGAGTGGAGGACCTGACCCAGCAGTACTACTCCCTCGTTTATGCGGATGCTTTC 470
DB 934 CCGAGGCGTGTGCTCTCTTCTCCAGCGCTACTCAAGTGTGTCCAGCAAGGGGAGAA 993
QY 471 ATCTCCAATTCAGAAATTTACTTTTTCGAGTCCAGGCTCTGCTGAATAGAACTGCTCT 530
DB 994 CTCAGGACCTCGACACTCAGGTTCAGGCTGTGAGGACATCTCTGACGAGCTGCAGGCC 1053
QY 531 GTGAACCTCTCAGACGACATTCAGCGGCTCCGGAACATCTGGATTATTAAGCCCGCGGCC 590
DB 1054 GTGGTACCCAGATAGACATGAGAGGGGATCGCNACATCTGGATCTGAGGACGAGNGCC 1113

Db 1583 AAGTCCCGCGGACGAGGCATCATGTGCATGGACCACTGGAGGAGATGCTGAAGCTGGT 1642
QY 651 GCTGCAGACCACTCTTTCCAGGACAAAGAGTGGTGGTCCAGAAGTACATCGACG 710
Db 1643 AACGGCAACCCGTTGGTATGAAGACGGCAAGTGGTGGTGCAGAAAGTATATTGACGG 1702
QY 711 CCGCTGCTCATCTGTGACACCAAGTTCGACATCAGACAGAGTGGTTCCTCGTCAGGACTGG 770
Db 1703 CCCCTCTCATCTTTGGACCAAGTTGACTCAGACAGTGGTTCCTGGTAACTGACTGG 1762
QY 771 AACCCCTGACCAATCTCTGTCTTACAAGGAGATTAATTGGCGTTCTCAACTCAGCGCTTC 830
Db 1763 AACCCACTTACCGTGTGTTCTACCGCGACAGCTATATCCGCTTTTCCAGCGACCCCTTC 1822
QY 831 TCCCTGACAGCTGGACAGGCGCATCCACTGTGCAACACGCGCTGCAGAACTACCTG 890
Db 1823 TCCCTGAGAACCTTGGCAACTCAGTCACTGACCTGTGCAACAACTCCATCCAGAACCTG 1882
QY 891 AAGAAATGATGTGGCCCGCAGCCCTGCTGCCGACACAAACATGTGGACAGCACCAGG 950
Db 1883 GAGAACTCATGCCATCGGCATCACTGCTTCCGCGACAGCAACATGTGGTCTAGCCAGG 1942
QY 951 TTCAGAGTACCTGACGCGCCAGGCGCGTGGCGCCGTGTGGGACGCGTCACTTACCCG 1010
Db 1943 TTCAGGCCCACTGACGAGAGTGGGTGCCCAAAATGCTTGGTCCACCATCATGTCGCT 2002
QY 1011 TCCATGAAGAGCCATCGCCAGCCCATGAAGTGGCCGACGACCACTGGAGCCTCGC 1070
Db 2003 GGCATGAAGATGCTGTGATCCAGCACTTCCAGACCTCCAGGACACCGTGCAGTGTGCG 2062
QY 1071 AAGAAAGCTTTGAGCTCTACGGGGCTGACTTCTGCTTGGAGGGACTTTCAGGCCCTGG 1130
Db 2063 AAGCCAGCTTTGAGCTCTATGGCGCTGACTTCTGTGTTGGGAGGACTTTCAGGCCCTGG 2122
QY 1131 CTGATCGAGATCAATTCAGCCCCACATGACCCCGTCCACGCGGTTCAGCGCCCAAGCTG 1190
Db 2123 CTGATTCAGATCAACGCCAGCCCCACGATGCGACCCCTCCACAGCAGTCACTGCGCGCTC 2182
QY 1191 TGTGCACAGTGCAGGAGGACACATCAAGTGGCC-----GTGACCGGAGC 1238
Db 2183 TGTGCTGGCTGCAAGCTGACACCTCGCGGTGTGCTATTGACCGGATGCTGGACCGCAAC 2242
QY 1239 TGTGACATCGCACTTCGAGCTCCTGTGGAGCGACCGGT 1279
Db 2243 TGTGACAGAGGCTTTGAGCTCATCTATAGCAGCCCGT 2283

RESULT 15

ADQ98824

ID ADQ98824 standard; cDNA; 3001 BP.

XX AC

XX ADQ98824;

XX DT 23-SEP-2004 (first entry)

XX DE

DE DNA encoding human GPCR-like protein seqid 494.

XX KW

KW ophthalmological; immunomodulatory; cytostatic; antiatherosclerotic;
KW antidiabetic; GPCR-like protein; ophthalmic disorder;
KW neurological disorder; immunological disorder; nephritic disorder;
KW hormonal dysfunction; cancer; atherosclerosis; diabetes;
KW molecular weight marker; food supplement; human; ss.

XX OS

XX Homo sapiens.

XX PN

XX US659662-B1.

XX XX

PD 27-MAY-2003.

XX XX

PF 19-JUL-2000; 2000US-00620312.

XX XX

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR

XX

PA

(HYSE-) HYSEQ INC.

XX

Tang YT, Zhou P, Drmanac RT;

XX

WPI; 2001-442255/47.

XX

New G-protein-coupled receptor-like polypeptides and polynucleotides,
useful for treating diseases of ophthalmic, neurological, immunological
and nephritic systems and hormonal dysfunction, cancer, atherosclerosis
and diabetes.

XX

Example 2; SEQ ID NO 494; 92pp; English.

XX

The invention describes an isolated polynucleotide (I) comprising a fully
defined (S1) of 749, 3188, 2484, 1169, 2936, 1467, 5773, 5714, 4041,
CC 1372, 3996, 3945, 2735, 1788, 585, 1782, 927, 5714 or 2282 nucleotides as
CC given in the specification, its translated or protein coding portion, its
CC extracellular portion or its active domain. The GPCR-like polypeptides
CC and polynucleotides are useful for the treatment of diseases of
CC ophthalmic, neurological, immunological and nephritic systems. They may
CC also be used to treat hormonal dysfunction, cancer, atherosclerosis and
CC diabetes. The antibodies are useful for detecting or quantitating the
CC polypeptide in tissue. The polypeptides can also be used as molecular
CC weight markers and as a food supplement. This sequence represents a human
CC polynucleotide of the invention.

XX

SQ Sequence 3001 BP; 662 A; 854 C; 846 G; 639 T; 0 U; 0 Other;

XX

Query Match 22.9%; Score 421.4; DB 5; Length 3001;

XX

Best Local Similarity 63.8%; Pred. No. 3.1e-80;

XX

Matches 677; Conservative 0; Mismatches 371; Indels 13; Gaps 2;

QY

232 GGGAGAGGCGGGAGCAGACCTGAGCAGCAGGCAAGATCTGAAATGCTGAGGCAA 291

Db

1223 GGGATCGTGACAGGCCCTCCCTATGTGAGGAGACAAGCAGCCCAAGAAACAGGAGAA 1282

QY

292 AGCTCAG-GGGCTCCCGGGGAGCTTGTGGACATCGCTGCAAGGTGTGCCAGGCTAC 350

Db

1283 AACCCAGTGTGGTGTCCCGAGAGTTTGTGGATGAAGCTCTGTGTGCTGCGAGGAGTAC 1342

QY

351 CTGGGCGAGCTGGAGCATGAGGACATCGACACCTCAGCAGATCCGCTGGAGGACCTCCTC 410

Db

1343 CTTAGCAACTTGGCCCACTGACATCGACAGGACCTGAGGCCCGCTGTACTCACC 1402

QY

411 GAGGCGAGTGGAGGACCTGACCCAGCAGTACTACTCCTCGTTTCATGCGCATGCTTTC 470

Db

1403 CCGAGGGCTGGTCCCTCTTCTCCAGCGCTACTACCAAGTGTCTCCAGAGGGGAGAA 1462

QY

471 ATCTCCAATTCAAGAAATTAATTTCTTTCGAGTGCCAGGCTCTGCTGAATAGATCAGTCT 530

Db

1463 CTCAGGCACCTCGACACTCAGGTCCAGCGCTGTGAGGACATCTCTGACGAGCTGCA 1522

QY

531 GTGAACCTTCAGACGACATGACGGGCTCCGAAACATCTGGATTATAAGAGCCCGGGCC 590

Db

1523 GTGTACCCAGATAGACATGGAAGGGATCGCAACATCTGGATCTGTAAGCCAGAGGCC 1582

QY

591 AAGTCCCGGGCGGACATAGTGTGATGACCGCTGTGGAGAGATCTCTGGAGCTGGCA 650

Db

1583 AAGTCCCGGGAGGAGCATGTCATGGACCACTCTGGAGGAGATGCTGAAGCTGGTG 1642

QY

651 GCTGACAGCACCTCTTTCAGGGAACAAGTGGTGGTCCAGAGTACATCGAGAGC 710

Db

1643 AACGGCAACCCGTTGTGTATGAAGGACGGCAAGTGGTGGTGCAGAGTATATTGAGCGG 1702

QY

711 CCGCTGCTCATCTGTGACACCAAGTTTCGACATCAGACAGTGGTTCCTCGTCAGGACTGG 770

Db

1703 CCCCTCTCATCTTGGCACCAAGTTTGACCTCAGACAGTGGTTCCTGGTAACTGACTGG 1762

QY

771 AACCCCTCAGCATCTGGTTCTACAGGAGAGTTACTTGGCGTTCTCAACTCAGCGCTTC 830

Db

1763 AACCCACTTACCGTGTGGTTCTACCGCGACAGCTATTCGCTTTTCCAGCGAGCCCTTC 1822


```
QY      831  TCCTGGACAGCTGACAGCGCCATCCACCTGTGCAACAAACGCCCTCCAGAAGTACCTG 890
Db      1823  TCCTGAAGAACCTGGACAACCTCAGTGCACCTGTGCAACAACTCCATCCAGAAGCACCTG 1882
QY      891  AAGAATGATGTGGCGCGCAGCCCTGCTGCGCGACACAAACATGTGGACCGACCCAGS 950
Db      1883  GAGAACTCATGTCATCGGCATCCACTGCTTCGCCACAGAACAAATGTGGTCTAGCCAGAGG 1942
QY      951  TTCAGGAGTACCTGCAGCGCCAGGCGCGTGGCGCGTGTGGGGCAGCGTCACTTACCCG 1010
Db      1943  TTCAGGCCCACTGTGAGGATGGGTGCCCCAAATGCTTGGTCCACCATCATCTGTCCT 2002
QY      1011  TCATGAAGAAGGCCATCGGCCACGCCATGAAGGTGGCCCGACGACCAACGTGAGCCTCGC 1070
Db      2003  GGCATGAAGGATGCTGTGATCCACGCACTTCAGACCTCCAGGACACCGTGCAGTGTGCG 2062
QY      1071  AAGAACAGCTTTGAGCTCTACGGGGTGAATTCGTCCTTTGGGAGGAGACTTCAGGCCCTGG 1130
Db      2063  AAGGCCAGCTTTGAGCTCTATGGCGTGAATTCGTCCTTCGGGAGGAGACTTCAGGCCCTGG 2122
QY      1131  CTGATCGAGATCAATTCAGCCCCACATGCACCCCGTCCACGCCGTACAGGCCCAAGCTG 1190
Db      2123  CTGATTGAGATCAACGCCAGCCCCACGATGGCACCCCTCCACAGCAGTCACTGCCCGGCTC 2182
QY      1191  TGTGCACAGGTGAGGAGGACACCATCAAGGTGGCC-----GTGACCGCAGC 1238
Db      2183  TGTGCTGGCGTGCAAGCTGACACCCCTGCGCGTGGTCAATTGACCGGATGCTGGACCGCAAC 2242
QY      1239  TGTGACATGGCAACTTCGAGCTCCTGTGGAGGCGCCGCT 1279
Db      2243  TGTGACACAGGAGCCTTTGAGCTCATCTATAAGCAGCCCGT 2283
```

Search completed: September 24, 2005, 18:03:24
Job time : 1125.21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 11:30:53 ; Search time 10402.7 Seconds
(without alignments)
10438.446 Million cell updates/sec

Title: US-10-615-659-27
Perfect score: 2241
Sequence: 1 tgcctcagacagcccgccg.....gcctccacagggccctccc 2241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	2241	100.0	101270	9 HS355C18	AL022327 Human DNA
2	136	6.1	374	6 CQ724916	CQ724916 Sequence
3	108.8	4.9	163783	10 AC119959	AC119959 Mus muscu
4	102.6	4.6	218249	2 AC097425	AC097425 Rattus no
5	102.2	4.6	125020	9 AF429315	AF429315 Homo sapi
6	101.6	4.5	773	11 BV054314	BV054314 S212P6049
C 7	79	3.5	125020	9 AF429315	AF429315 Homo sapi
8	76.6	3.4	197360	2 AC018829	AC018829 Homo sapi
9	76.4	3.4	145435	2 AC026685	AC026685 Homo sapi
C 10	76.4	3.4	185067	9 AC022382	AC022382 Homo sapi
C 11	76.4	3.4	189430	2 AC011610	AC011610 Homo sapi
12	76.4	3.4	191834	2 AC026196	AC026196 Homo sapi
C 13	76.4	3.4	216911	9 AC018809	AC018809 Homo sapi
14	76.4	3.4	222994	2 AC016947	AC016947 Homo sapi
15	73.2	3.3	155313	2 AC068315	AC068315 Homo sapi
16	72.4	3.2	1299	10 BC021404	BC021404 Mus muscu
17	71.6	3.2	5282	6 CQ842940	CQ842940 Sequence
18	71.6	3.2	5282	9 AK125875	AK125875 Homo sapi
C 19	71.6	3.2	224077	2 AC120010	AC120010 Mus muscu

C	20	71.6	3.2	266888	2	AC137881	AC137881 Mus muscu
C	21	71.4	3.2	252689	2	AC079433	AC079433 Mus muscu
C	22	69.8	3.1	7218	6	166494	166494 Sequence 14
C	23	68.6	3.1	209887	2	AC079424	AC079424 Mus muscu
C	24	68.2	3.0	1897	10	BC006830	BC006830 Mus muscu
C	25	67	3.0	64976	2	AC026256	AC026256 Homo sapi
C	26	67	3.0	267894	2	AC117950	AC117950 Rattus no
C	27	66.8	3.0	239130	2	AC079420	AC079420 Mus muscu
C	28	66.4	3.0	202083	2	AC023833	AC023833 Mus muscu
C	29	65.4	2.9	300695	2	AC079431	AC079431 Mus muscu
C	30	65	2.9	303091	2	AC084799	AC084799 Mus muscu
C	31	64	2.9	126543	2	AC143806	AC143806 Macaca mu
C	32	63.8	2.8	265537	2	AC087228	AC087228 Mus muscu
C	33	63.4	2.8	3281	2	AC143602	AC143602 Macaca mu
C	34	63	2.8	261449	2	AC095574	AC095574 Rattus no
C	35	62.6	2.8	63087	2	AC023446	AC023446 Homo sapi
C	36	62.4	2.8	736	11	BV106786	BV106786 PZA01640
C	37	62.4	2.8	840	8	CNS01BNS	AL114464 Botrytis
C	38	62.4	2.8	134940	2	AC018939	AC018939 Homo sapi
C	39	62.2	2.8	141003	2	AC016463	AC016463 Homo sapi
C	40	62.2	2.8	161903	2	AC021996	AC021996 Homo sapi
C	41	62	2.8	250046	2	AC113592	AC113592 Mus muscu
C	42	62	2.8	298166	2	AC087563	AC087563 Homo sapi
C	43	61.8	2.8	132001	2	AC142984	AC142984 Macaca mu
C	44	61.6	2.7	128658	2	AC147014	AC147014 Medicago
C	45	61.6	2.7	207420	2	AC078884	AC078884 Mus muscu

ALIGNMENTS

RESULT 1
HS355C18/c

LOCUS

DEFINITION

HS355C18 101270 bp DNA linear PRI 05-JUN-2003
Human DNA sequence from clone RP3-355C18 on chromosome 22q13.3
Contains the KIAA0027 gene, ESTs, STSs, GSSs and seven putative CpG islands, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL022327

AL022327.17 GI:5304851

HTG; CpG island; KIAA0027.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 101270)

Cobley, V.

Direct Submission

Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquiesanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Jun 30, 1999 this sequence version replaced gi:5262834.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submissions with

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquiesanger.ac.uk

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed from the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>
RP3-355C18 is from the library RPCI-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-355C18 The true left end of clone RP3-898I4 is at 5390 in this sequence.

FEATURES

source	Location/Qualifiers	
	1. .101270	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosomes="22"	
	/map="q13.3"	
	/clone="RP3-355C18"	
	/clone_lib="RPCI-3"	
misc_feature	1. .1038	
	/note="match: STS: Em:AL022483"	
repeat_region	complement(1. .242)	
	/notes="AluJo repeat: matches 1. .237 of consensus"	
repeat_region	358. .484	
	/note="L1M4 repeat: matches 829. .959 of consensus"	
repeat_region	485. .792	
	/notes="AluSx repeat: matches 2. .305 of consensus"	
repeat_region	795. .1095	
	/notes="AluSx repeat: matches 1. .299 of consensus"	
repeat_region	1096. .1165	
	/note="L1M4 repeat: matches 960. .1031 of consensus"	
repeat_region	1166. .1563	
	/notes="MSTB1 repeat: matches 4. .432 of consensus"	
repeat_region	1564. .1750	
	/note="L1M4 repeat: matches 1031. .1203 of consensus"	
repeat_region	1757. .2038	
	/note="AluSx repeat: matches 1. .282 of consensus"	
repeat_region	complement(2054. .3342)	
	/notes="SVA repeat: matches 1. .1386 of consensus"	
repeat_region	3343. .3359	
	/notes="2.8 copies 6 mer TCTCTT 34% conserved"	
repeat_region	3360. .3417	
	/note="LNR39-int repeat: matches 3904. .3961 of consensus"	
repeat_region	3418. .3577	
	/note="L1M4 repeat: matches 1203. .1363 of consensus"	
repeat_region	3578. .3861	
	/note="AluJo repeat: matches 1. .300 of consensus"	
repeat_region	3862. .4334	
	/note="L1M4 repeat: matches 1363. .1830 of consensus"	
repeat_region	4341. .4489	
	/note="L1M4 repeat: matches 2065. .2311 of consensus"	
repeat_region	complement(4492. .4840)	
	/notes="MER4A1 repeat: matches 96. .472 of consensus"	
repeat_region	4990. .4999	
	/note="2.5 copies 4 mer CCTG 20% conserved"	
repeat_region	complement(5142. .5300)	
	/note="AluJb repeat: matches 137. .295 of consensus"	
repeat_region	complement(5315. .5623)	
	/notes="AluSx repeat: matches 3. .312 of consensus"	
repeat_region	5773. .5863	
	/note="MER81 repeat: matches 17. .112 of consensus"	
repeat_region	5917. .5935	
	/notes="2.7 copies 7 mer GGGCCA 29% conserved"	
repeat_region	6130. .6140	
	/notes="3.7 copies 3 mer GGT 22% conserved"	
repeat_region	6139. .6148	
	/note="2.5 copies 4 mer GGAG 20% conserved"	
repeat_region	6229. .6240	
	/notes="2.0 copies 6 mer CAGCAC 24% conserved"	
repeat_region	6573. .6590	

repeat_region	6691. .6700	/note="3.0 copies 6 mer CAGGAG 27% conserved"
	/note="2.0 copies 5 mer TGTGC 20% conserved"	
repeat_region	6731. .6862	
	/note="3.3 copies 40 mer	
	CCAGGTTAGGAGGACCTGGGAGTGGCAGCTGCA 219% conserved"	
	7338. .7356	
	/note="2.7 copies 7 mer GGGCCAC 29% conserved"	
repeat_region	7636. .7664	
	/note="2.1 copies 14 mer CCCAGCAGGTGCC 49% conserved"	
repeat_region	7893. .7906	
	/note="2.0 copies 7 mer TGCACCC 28% conserved"	
repeat_region	7969. .7980	
	/note="2.0 copies 6 mer CCATGA 24% conserved"	
repeat_region	8306. .8316	
	/note="2.8 copies 4 mer GTGG 22% conserved"	
repeat_region	8399. .8414	
	/note="3.2 copies 5 mer GAGGG 23% conserved"	
repeat_region	8840. .8851	
	/note="2.4 copies 5 mer AGCTG 24% conserved"	
repeat_region	8939. .8958	
	/note="2.2 copies 9 mer CCCCTCCCA 40% conserved"	
repeat_region	9125. .9135	
	/note="2.2 copies 5 mer GTCCT 22% conserved"	
repeat_region	9246. .9256	
	/note="3.7 copies 3 mer GGA 22% conserved"	
repeat_region	9263. .9274	
	/note="2.0 copies 6 mer CAGTGA 24% conserved"	
repeat_region	9373. .9383	
	/note="2.2 copies 5 mer CCTGA 22% conserved"	
repeat_region	9434. .9444	
	/note="2.2 copies 5 mer CACGC 22% conserved"	
repeat_region	9616. .9631	
	/note="2.3 copies 7 mer GGGGATG 32% conserved"	
repeat_region	10067. .10076	
	/note="2.5 copies 4 mer GGCA 20% conserved"	
repeat_region	10245. .10254	
	/note="2.5 copies 4 mer CCTT 20% conserved"	
repeat_region	10496. .10531	
	/note="2.0 copies 18 mer GGCAGGACAGGACAGG 63% conserved"	
repeat_region	10498. .10531	
	/note="2.8 copies 12 mer CAGGGCAGGGA 59% conserved"	
repeat_region	10745. .10756	
	/note="2.0 copies 6 mer GCTGGG 24% conserved"	
repeat_region	10889. .10903	
	/note="2.1 copies 7 mer CCGGCC 30% conserved"	
repeat_region	11165. .11193	
	/note="3.5 copies 8 mer TGTGTGCA 33% conserved"	
repeat_region	11177. .11211	
	/note="5.8 copies 6 mer GTGCAT 38% conserved"	
repeat_region	11250. .11261	
	/note="2.0 copies 6 mer ATGGGT 24% conserved"	
repeat_region	11273. .11282	
	/note="2.5 copies 4 mer TGCC 20% conserved"	
repeat_region	11293. .11309	
	/note="2.8 copies 6 mer TGTGTA 25% conserved"	
repeat_region	11311. .11411	
	/note="2.5 copies 40 mer	
	TGCAGGTTAGTATCGGTATGGTGTGCTGAGTGCCT 202% conserved"	
repeat_region	11323. .11334	
	/note="2.0 copies 6 mer ATGGGT 24% conserved"	
repeat_region	11346. .11355	
	/note="2.5 copies 4 mer TGCC 20% conserved"	
repeat_region	11368. .11379	
	/note="2.0 copies 6 mer ATGGGT 24% conserved"	
repeat_region	11391. .11400	
	/note="2.5 copies 4 mer TGCC 20% conserved"	
repeat_region	11431. .11683	
	/note="7.0 copies 36 mer	
	TGCATGTATGGTGTGTATGTGATGGTGGTGGTG 108% conserved"	
repeat_region	11451. .11490	
	/note="2.9 copies 14 mer TGTGCATGGGTGG 71% conserved"	
repeat_region	11452. .11477	

		/note="3.2 copies 8 mer GTGCATGG 29% conserved" 11485..11498 /notes="2.3 copies 6 mer TGGGTG 28% conserved" 11513..11523 /note="2.8 copies 4 mer TGCC 22% conserved" 11521 /note="Tandem repeat. Assembly not confirmed by digest" 11531..11562 /notes="5.3 copies 6 mer TGTGCA 32% conserved" 11536..11635 /notes="6.2 copies 16 mer GTGTGCATGTGTATGG 82% conserved" 11595..11604 /notes="5.0 copies 2 mer TG 20% conserved" 11679..11690
		Query Match Best Local Similarity 100.0%; Score 2241; DB 9; Length 101270; Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	TGCTCGAGCAGAGCCGGCTGGATCTTCACGCTGCACGCCATTGCTCTCTCATGGTT 60
Db	60138	TGCTCGCAGCAGAGCCGGCTGGATCTTCAGCTGCAGCCATTGCTCTCTCATGGTT 60079
Qy	61	GCTCACTCCATCCCATCTGGCTGTGTGAACCTTGAATGCTGGGCATCAATAAGACT 120
Db	60078	GCCTCACTCCATCCCATCTGGCTGTGTGAACCTTGAATGCTGGGCATCAATAAGACT 60019
Qy	121	TTTTTCTTGCCAGCTTAGGTTGGTAAATAAGATGCGTACCCTCTCTCGAACCATCAA 180
Db	60018	TTTTTCTTGCCAGCTTAGGTTGGTAAATAAGATGCGTACCCTCTCTCGAACCATCAA 59595
Qy	181	GAGGACATCATTTGACTATCACAGCCTGACCTACGACACAGATGCTGAACCACTACGCAAA 240
Db	59958	GAGGACATCATTTGACTATCACAGCCTGACCTACGACACAGATGCTGAACCACTACGCAAA 59899
Qy	241	GACAGCCTCTTTCACCAACGAGTGACCGCGCGCCTTGAAGGTGCGGAGACCGGCAA 300
Db	59898	GACAGCCTCTTTCACCAACGAGTGACCGCGCGCCTTGAAGGTGCGGAGACCGGCAA 59839
Qy	301	GGGTGTGGGGAGGAGCTTCTGCAGACAATTAGGGACCTTCGGTGGCTGTGATGCGCCAC 360
Db	59838	GGGTGTGGGGAGGAGCTTCTGCAGACAATTAGGGACCTTCGGTGGCTGTGATGCGCCAC 59779
Qy	361	CAGTCACCTCTCTGGGTGCATCCAGGACTCTCTGGTCTAAGGCCGTGGCCAGAATCACTCG 420
Db	59778	CAGTCACCTCTCTGGGTGCATCCAGGACTCTCTGGTCTAAGGCCGTGGCCAGAATCACTCG 59719
Qy	421	GTGCCACCCACCCACCCAGCACCCCTGTGCGCTTTGCTCTGTGTCTCTGGGTGAATCCGG 480
Db	59718	GTGCCACCCACCCACCCAGCACCCCTGTGCGCTTTGCTCTGTGTCTCTGGGTGAATCCGG 59659
Qy	481	GGCCCCAGAGCTCCCTCTCAGGGCACAGAGCCAAAAGATGGGCTGACTGGGGGGCTGC 540
Db	59658	GGCCCCAGAGCTCCCTCTCAGGGCACAGAGCCAAAAGATGGGCTGACTGGGGGGCTGC 59599
Qy	541	CACCGGGCTTTGGGTGCTGAGGGGGCTGTGGGACCCACAGGGAAAGAGTGCCTGCCCTCC 600
Db	59598	CACCGGGCTTTGGGTGCTGAGGGGGCTGTGGGACCCACAGGGAAAGAGTGCCTGCCCTCC 59539
Qy	601	CAGCACGGGCACTCCCCCGCTCTCCCCCGCTCTCAGACCTTCAGCAAAAGCAGCTTGCAC 660
Db	59538	CAGCACGGGCACTCCCCCGCTCTCCCCCGCTCTCAGACCTTCAGCAAAAGCAGCTTGCAC 59479
Qy	661	ACCTTTCCACCAAGGCCACAGGCTGAGGCTGCCACGACCGAGGTGTGGGGACCTTG 720
Db	59478	ACCTTTCCACCAAGGCCACAGGCTGAGGCTGCCACGACCGAGGTGTGGGGACCTTG 59419
Qy	721	CTGAGGAGGGGTCCCGGCAAGGGGCTTCCTCTGGCACAGAGGTCCCTCTCAGCAGGCGCA 780
Db	59418	CTGAGGAGGGGTCCCGGCAAGGGGCTTCCTCTGGCACAGAGGTCCCTCTCAGCAGGCGCA 59359
Qy	781	GGTGGCGCTGCCTCAGCA CAGTGTGGGGCGGAGGTGCAGCA CAAAGTTCCCTTCCGCAAC 840

59358	Db	GGTGGCGCTGCCCTCAGACAGTGTGGGGCGAGGTGCAGGA	CAAGGTTCCCTTCCGTCACCC	59299
841	Qy	TAATAACCCCAAGGTCAGCGCCAGCGCCTAGTGGCAACATGGCCCTTCAAAAGA	900	
59298	Db	TAATAACCCCAAGGTCAGGCCAGCGCTAGTGGCAACATGGCCCTTCAAAAGA	59239	
901	Qy	CCCGTGTGCAGAGCCAGTCAGCGTGGCTGTGTCTCTGGGGGCCAGCCACTTTTTC	960	
59238	Db	CCCGTGTGCAGAGCCAGTCAGCGTGGCTGTGTCTCTGGGGGCCAGCCACTTTTTC	59179	
961	Qy	TTTAACAGGTTGACAAACACAGACATCGGGGTGGGGACCTTCAAGATGCGCTCTCTAATG	1020	
59178	Db	TTTAACAGGTTGACAAACACAGACATCGGGGTGGGGACCTTCAAGATGCGCTCTCTAATG	59119	
1021	Qy	GCGCCTGCTACCTCGCGCTGTGGTGGGCAGTCGTGGGCATGTGGCTGCGTTCGGGGCCCTG	1080	
59118	Db	GCGCCTGCTACCTCGCGCTGTGGTGGGCAGTCGTGGGCATGTGGCTGCGTTCGGGGCCCTG	59058	
1081	Qy	CAGCGCTCTCCCTCTCTCTGGGGTCTGAGGTGTGGGAACACCTGTGGGCCGTGTGGGG	1144	
59058	Db	CAGCGCTCTCCCTCTCTCTGGGGTCTGAGGTGTGGGAACACCTGTGGGCCGTGTGGGG	58999	
1141	Qy	CCTGAGGAGGAGCCCAAGTGCACCGCCCTCGGTGGCCCTCAGCTGCAGGCTCCCAATTCC	1200	
58998	Db	CCTGAGGAGGAGCCCAAGTGCACCGCCCTCGGTGGCCCTCAGCTGCAGGCTCCCAATTCC	58939	
1201	Qy	CTGGGGCCTGCCACGCCCTCCACTGCTCAAGGATGCGTCTCAGGCTGGCCTGCACATCC	1260	
58938	Db	CTGGGGCCTGCCACGCCCTCCACTGCTCAAGGATGCGTCTCAGGCTGGCCTGCACATCC	58879	
1261	Qy	ATGCTCCCAACCGGCTCGAGGGGGTGTCTGGGGCCCTGGGGACCGTATATCTGGGGAGG	1320	
58878	Db	ATGCTCCCAACCGGCTCGAGGGGGTGTCTGGGGCCCTGGGGACCGTATATCTGGGGAGG	58819	
1321	Qy	GGAGGGATCCGCCCTAGGGGTGCTGTGTGGTGGCCATAGCCAGGCAAGTCTGAGGTGG	1380	
58818	Db	GGAGGGATCCGCCCTAGGGGTGCTGTGTGGTGGCCATAGCCAGGCAAGTCTGAGGTGG	58759	
1381	Qy	ACGGCCGCTCTCGCTGTGAGCAGGAGCAGAGCTGAGCTGAGCTGCCCTCTCGCTAA	1440	
58758	Db	ACGGCCGCTCTCGCTGTGAGCAGGAGCAGAGCTGAGCTGAGCTGCCCTCTCGCTAA	58699	
1441	Qy	CACGCCACTGTCTCTGGAGATCGGGCTGTGCGTGAACATGCGGAGCTGCCCTCGCTAA	1500	
58698	Db	CACGCCACTGTCTCTGGAGATCGGGCTGTGCGTGAACATGCGGAGCTGCCCTCGCTAA	58639	
1501	Qy	CCCGGCCAACCCCGACTCTTCTTCCCAAGCTCTACAGCTCTGCAACGAGAGTACGA	1560	
58638	Db	CCCGGCCAACCCCGACTCTTCTTCCCAAGCTCTACAGCTCTGCAACGAGAGTACGA	58579	
1561	Qy	GCAGGAGTTCTCGGTAACTGAGGAGACGGCAGAGGGCCCAAGTGTGTGCACACAGAGG	1620	
58578	Db	GCAGGAGTTCTCGGTAACTGAGGAGACGGCAGAGGGCCCAAGTGTGTGCACACAGAGG	58519	
1621	Qy	CTTCTAGAAAGATCCCCCTGGTGTGAGACAGACTGATGGGGCAGGGTCTGAGGATAGAG	1680	
58518	Db	CTTCTAGAAAGATCCCCCTGGTGTGAGACAGACTGATGGGGCAGGGTCTGAGGATAGAG	58459	
1681	Qy	GACCGGGGAGAGCCTCCATGGTCAATGGTCAATGGCAGTACAGAGCCAGGGCCCCCGG	1740	
58458	Db	GACCGGGGAGAGCCTCCATGGTCAATGGTCAATGGCAGTACAGAGCCAGGGCCCCCGG	58399	
1741	Qy	AGGAGGGGGGAGGAGTGTGTGTCTGTGTAGGCTTTGTGACAGAGCCGCGG	1800	
58398	Db	AGGAGGGGGGAGGAGTGTGTGTCTGTGTAGGCTTTGTGACAGAGCCGCGG	58339	
1801	Qy	AGCAGCTGGGGCAAGGCCCGGCTCCGTGTGCGAGGCGCGAGGGGTGCAGTCGAGCCGCG	1860	
58338	Db	AGCAGCTGGGGCAAGGCCCGGCTCCGTGTGCGAGGCGCGAGGGGTGCAGTCGAGCCGCG	58279	
1861	Qy	AGTCACAGACAGACTGCAGGGAGAGGAGGAGGAGACTGTGGGGTGTGGGAGCAGCCCTGG	1920	
58278	Db	AGTCACAGACAGACTGCAGGGAGAGGAGGAGGAGACTGTGGGGTGTGGGAGCAGCCCTGG	58219	

Qy	1921	GC	CCCATCGCCTATAGGGCGGGAAGTGGGGCGGGTTTGGGGAGCCTCCGTGGCCCTGGCC	1980
Db	58218	GC	CCCATCGCCTATAGGGCGGGAAGTGGGGCGGGTTTGGGGAGCCTCCGTGGCCCTGGCC	58159
Qy	1981	CG	ATCTCCAGTCCCGAGTCTCTGGCTCGGACAGATAGGGCGAGGCTGTGCTGCTCTTCAG	2040
Db	58158	CG	ATCTCCAGTCCCGAGTCTCTGGCTCGGACAGATAGGGCGAGGCTGTGCTGCTCTTCAG	58099
Qy	2041	AG	CACTTCGGCGCCACCATGGCATCCAGCATCTCAAGTGGGTGTCAGCCACGAGCT	2100
Db	58098	AG	CACTTCGGCGCCACCATGGCATCCAGCATCTCAAGTGGGTGTCAGCCACGAGCT	58039
Qy	2101	GC	AGCAGGAGCAGCAGAGAAGCCAGAGCCAGGACACAGAGGAGGAGGCGCGGAGCAGCGACC	2160
Db	58038	GC	AGCAGGAGCAGCAGAGAAGCCAGAGCCAGGACACAGAGGAGGAGGCGCGGAGCAGCGACC	57979
Qy	2161	TG	AGCAGCAGGCAAGGTGCGCTGGCGGGCGGACAGGCGAGGCGAGTGCACGAGCCACGAGCTCGG	2220
Db	57978	TG	AGCAGCAGGCAAGGTGCGCTGGCGGGCGGACAGGCGAGGCGAGTGCACGAGCCACGAGCTCGG	57919
Qy	2221	GG	CTCCACAGGGGGCCTCCC	2241
Db	57918	GG	CTCCACAGGGGGCCTCCC	57898
RESULT 2				
LOCUS	CQ724916		374 bp	DNA linear PAT 03-FEB-2004
DEFINITION	Sequence 10850 from Patent WO02068579.			
ACCESSION	CQ724916			
VERSION	CQ724916.1		GI:42285773	
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
JOURNAL	1			
FEATURES	Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.			
source	Kits, such as nucleic acid arrays, comprising a majority of			
ORIGIN	humanexons or transcripts, for detecting expression and other uses			
	thereof			
	Patent: WO 02068579-A 10850 06-SEP-2002;			
	PE Corporation (NY) (US)			
	Location/Qualifiers			
	1..374			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			
Query Match	6.1%	Score 136;	DB 6;	Length 374;
Best Local Similarity	100.0%;	Pred. No. 2.8e-16;		
Matches 136;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	2040	GA	AGACTTCGCGCGCACCATGGCATCTCAAGTGGGTGTCAGCCACGAGC	2099
Db	239	GA	AGACTTCGCGCGCACCATGGCATCTCAAGTGGGTGTCAGCCACGAGC	298
Qy	2100	TG	CAGCAGGAGCGCAGAAAGCCAGGACAGAGGAGGAGGCGCGGAGCAGCGAC	2159
Db	299	TG	CAGCAGGAGCGCAGAAAGCCAGGACAGAGGAGGAGGCGCGGAGCAGCGAC	358
Qy	2160	CT	CAGCAGCGAGGCAAG	2175
Db	359	CT	CAGCAGCGAGGCAAG	374
RESULT 3				
LOCUS	AC119959			
DEFINITION	163783 bp	DNA linear	ROD 23-JUL-2004	
ACCESSION	AC119959			

AC119959.8	GI:50540805			
HTG.				
Mus musculus	(house mouse)			
Mus musculus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
1 (bases 1 to 163783)				
Birren, B., Nusbaum, C. and Lander, E.				
Mus musculus chromosome 15, clone RP24-467H19				
Unpublished				
2 (bases 1 to 163783)				
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,				
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,				
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,				
Chazaro, B., Choepel, Y., Collange, M., Collins, S., Collamore, A.,				
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,				
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,				
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,				
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,				
Kamat, A., Karatas, A., Kells, C., Laocque, K., Lamazares, R.,				
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,				
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,				
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,				
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,				
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,				
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,				
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,				
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,				
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,				
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,				
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,				
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,				
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.				
Direct Submission				
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome				
Research, 320 Charles Street, Cambridge, MA 02141, USA				
3 (bases 1 to 163783)				
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,				
Anderson, M., Anderson, S., Atachchi, H.M., Barna, N., Bastien, V.,				
Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J.,				
Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B.,				
Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,				
Ericksen, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,				
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,				
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,				
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,				
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,				
Maclean, C., Macdonald, P., Major, J., Manning, J., Meneus, L.,				
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,				
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,				
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,				
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,				
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,				
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,				
Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J.,				
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R.,				
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,				
Zimmer, A. and Zody, M.				
Direct Submission				
Submitted (26-MAY-2004) Whitehead Institute/MIT Center for Genome				
Research, 320 Charles Street, Cambridge, MA 02141, USA				
4 (bases 1 to 163783)				
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,				
Anderson, M., Anderson, S., Atachchi, H.M., Barna, N., Bastien, V.,				
Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J.,				
Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B.,				
Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,				
Ericksen, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,				
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,				
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,				
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,				
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,				
Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,				

McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
 Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
 Totham, K., Travers, M., Vasiliev, H., Venkataranan, V. S., Viel, R.,
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
 Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (23-JUL-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 23, 2004 this sequence version replaced gi:47679219.

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@broad.mit.edu

----- Project Information

Center project name: L25717

Center clone name: 467_H_19

FEATURES

source

Location/Qualifiers

1. .163783
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="15"
 /map="15"
 /clone="RP24-467H19"
 /clone_lib="RPCI-24 Male Mouse BAC"

misc_feature

1. .4
 /note="clone boundary
 clone end:SP6
 site:MboI"

repeat_region

complement (169. .204)
 /rpt_family="ORR1B1"
 818. .994
 /rpt_family="ORR1D"
 1312. .1527
 /rpt_family="B3"
 1573. .1645
 /rpt_family="ID3"
 1665. .1793
 /rpt_family="B1F"
 1794. .1959
 /rpt_family="B3"
 complement (2256. .2658)
 /rpt_family="L1"
 complement (2659. .2833)
 /rpt_family="B2_Mm2"
 complement (2834. .3302)
 /rpt_family="L1"
 complement (3311. .3448)
 /rpt_family="B1_MM"
 complement (3449. .4218)
 /rpt_family="L1"
 4214. .4557
 /rpt_family="Lx6"
 5120. .5279
 /rpt_family="CT-rich"
 complement (5297. .5423)
 /rpt_family="B1F"
 complement (5451. .5524)
 /rpt_family="PB1D9"
 6016. .6494
 /rpt_family="BLTR11A"
 complement (6678. .6797)
 /rpt_family="B1_MM"

repeat_region complement (6810. .6945)
 /rpt_family="B1_MM"
 7008. .7148
 /rpt_family="B1_MM"
 repeat_region complement (7763. .7941)
 /rpt_family="B3"
 repeat_region complement (7960. .8105)
 /rpt_family="B1_MM"
 repeat_region complement (8121. .8300)
 /rpt_family="B3A"
 8486. .8625
 /rpt_family="B4A"
 8618. .8788
 /rpt_family="B1_MM"
 repeat_region 8804. .9102
 /rpt_family="L1MB2"
 9162. .9217
 /rpt_family="PB1D7"
 repeat_region complement (9339. .9460)
 /rpt_family="B3"
 repeat_region complement (9598. .9748)
 /rpt_family="B3"
 repeat_region complement (9755. .12502)
 /rpt_family="L1_MM"
 12503. .12992
 /rpt_family="IAPLTR2_MM"
 12994. .13717
 /rpt_family="IAPLTR2_MM-int"
 13718. .13836
 /rpt_family="IAPLTR2_MM-int"
 13832. .14083
 /rpt_family="IAPLTR2_MM-int"
 14083. .15904
 /rpt_family="IAPLTR2_MM-int"
 15905. .16394
 /rpt_family="IAPLTR2_MM"
 complement (16395. .20012)
 /rpt_family="L1"
 repeat_region complement (20013. .20157)
 /rpt_family="B1_MM"
 20192. .20290
 /rpt_family="B2_Mm2"
 20410. .20553
 /rpt_family="B1_MM"
 20586. .20777
 /rpt_family="B2_Mm2"
 21012. .21082
 /rpt_family="(CCAA)n"
 21192. .21294
 /rpt_family="PB1D10"
 21301. .21342

Query Match 4.9%; Score 108.8; DB 10; Length 163783;

Best Local Similarity 84.7%; Pred. No. 1.9e-11;

Matches 122; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 122 TTTTCTTCCAGTCTAGGTTGGTAAAAATGAGATGCCGTACCTCCTCTGACCATCAAG 181
 Db 118788 TCTTTTCCAGTCCAGGTTGGTAAAGACGAGATCCATACCTCCTCTGACTATCAAG 118847
 Qy 182 AGGAGATCATTTGACTATCATCAGCCTGACCTACGACCATGCTGACCATCAAG 241
 Db 118848 AGGAGATTTGTGGACTACCAAGCTGACCTGACCATGCTGACCATCAAG 118907
 Qy 242 ACAGCTCTCTTCCACCAAGAGGTG 265
 Db 118908 ACAGATCATTCACCAAGAGGTG 118931

RESULT 4

AC097425

LOCUS

DEFINITION

AC097425 218249 bp DNA linear HTG 10-MAY-2003
 Rattus norvegicus clone CH230-137J2, *** SEQUENCING IN PROGRESS

LOCUS AF429315 125020 bp DNA linear PRI 18-JAN-2002
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S.,
Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A.,
Potter, N.T., Ross, C.A. and Margolis, R.L.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE
2 (bases 1 to 125020)
Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.
Direct Submission
TITLE Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
LOCATION/Qualifiers
1. 125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
repeat_region complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
gene complement(<36507..>36887)
/gene="JPH3"
mRNA /note="synonym: JP3"
complement(<36507..>36887)
/gene="JPH3"
product="junctophilin 3"
complement(<36507..36887)
/gene="JPH3"
CDS /note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAI40941.1"
/db_xref="GI:17646245"
translation="MSSGGRFNFDGSGYCGWEDGKAHGVCITGPKGGEYTGWS
HGFEVLGVYTPSGNTYGGTWAQGRHIGLESKGVYKGEWTHGPKRGYVRECA
NGAYEGTWSNGLQDGYGTETSDG"
ORIGIN
Query Match 4.6%; Score 102.2; DB 9; Length 125020;
Best Local Similarity 11.7%; Pred. No. 4e-10;
Matches 188; Conservative 711; Mismatches 689; Indels 16; Gaps 6;
QY 388 CTCCTGCTGAAGCGCTGGCCAGAACTCACTCGGTGCCCCACCCAGCCAGCACCCCTG 447
DB 50557 CTGCTGCTGGGCGCCACACCTCTGRTGKCNWSKSHSSCHRRPTWTDYRKYTYCCCYT 50616
QY 448 TGCCCTTTGCTCTGCTCTGCTGGAATCCGGGGCCCCAGAGCTCCCTCTCAGGGCA 507
DB 50617 AYMMBBYHYCYDBCWSGVWGSYGWGRGMKMBWBMBBMMGKSMGDTGDKDYVCVYSSM 50676
QY 508 CAGAGGCCAAGATGGGGCTGACTGGGGCTGCCACCGGGCTTTGGGTCTGAGGGGCT 567
DB 50677 SVDVYVWRWBDTSWGSBSHYMRVHTGVMRVWVGBCDTHVTYDGHGHSNMGKSCYS 50736
QY 568 GTGGACCCAGGGGAGAGGTGCGCTCCCGCCAGCACCGGGGCACTCCCGCGCTCCC 627
DB 50737 DBGKAKBGSXMTGSMYKTCYGYKWSKSDBSYYSBSHYBWSGMYMCCRKESYAS 50796

QY 628 CGGCTCAGCCCTGCACAAAGCAGCTTGCCACACCCCTTCCACCAAGGCCAGGGCTGA 687
DB 50797 SYDYYTKBTGMYRYVRSCSWYYKYBSMSVARSKSGKHYRBSMKSRKDMSTSSMW 50856
QY 688 GGCCTGCCAGCAGCGAGGCTGTGGGGACCTGCTGAGGGAGGGGTCCCGAAGAGGGGC 747
DB 50857 RSMKDRKCSRSHSHYSKSMGKWTSDVRCVACSBMYKWHKHYKDKDSHBSWGKSHM 50916
QY 748 TTCCTTGCCACAGAGG---TCCCTCTCAGAGGCGAGGTGCGGCTCGCTCAGCACAGT 804
DB 50917 KKGAWRVRSMYRSMKCSVHDSMVDKSRMRKSGAMRMKCYCTSSMRSTKSRSMKSSW 50976
QY 805 GGGG---CGAGGTGAGGACAAGTTCCTTCGCGACCTTAATACCCAGGGTCAGGCC 860
DB 50977 RWGSKYCYGYWSTCKMRSMGCSYSTGSSMKCMGKYCAGAGRSSRYCKSGSRMSAG 51036
QY 861 AGCCGACGCTGCTAGTGCGCAACATGGCCCTCTCAAGAGACCCCGTGTGCAGAGCAGTC 920
DB 51037 GMSASMSMGKMTSGMSRCRCARSRMSAYSCRKCKSKRCGCMRCSWYCWYKMRVGS 51096
QY 921 AGCGTCGCTGTGTTTCTGGGGCCAGCCACTTTTTCCTTAACAGGGTGACAAACAG 980
DB 51097 WRMYKCMWSYKTCYKGCATKSMAMYYVSWMRKGTKSSAWYSMCCSKCARAASGYK 51156
QY 981 ACATCGGGGTGCGGGAGCTTCAAGATGCTCTCTAATGGCGGTGCTACTCCGCTGT 1040
DB 51157 CMMKGGCYTGGGKGSASTWYMMCCSRAC--KYTTTRKSSRGRWGKATRTTRTSSRM 51214
QY 1041 GGTGGGAGTGTGGGATGCTGGCTTCGGGGCCCTGCGCGCTCCCTCCCTCTCTCT 1100
DB 51215 SWTYTAMMRRTWTTAGKWKTKTTWTKKSKMRMTRSRWNGMTTWTYTKTTTDDV 51274
QY 1101 GGGTCTGAGGTGTGGGACACCCCTGTGGCGGCTTGGGGGCTCGAGGAGCCCGCAGTC 1160
DB 51275 YKKKSKMKSKWYTTCTKYVYKTYKYTYKTYTGAGGYKRWGKCTCRMWSRWKWR 51334
QY 1161 CCACCGCCCTCGGTGCGCTCAGCTGAGCTGCCCATTTCCCTCGGGGCTGCGACGCCCTC 1220
DB 51335 GCMGAGGGSWRGAKYSMMSTSKYGRWTQWRSCTWCTSCMTRSCCTCMGCTCCTSMS 51394
QY 1221 CACTGCTCAAGGATGCTCTGAGCTGGCTGCACATCCATGCTCCACCGGCTGGAGG 1280
DB 51395 KRGTYSRRCYAVWSKCSYSYCMRCCTYSYMSKRSCTRRKWTYSWGRYKYCKRMSW 51454
QY 1281 GGGTGTCTCGGGCCCTGGGGGACCGTATATCTGGGAGGGGAGGATCCGCCCTAGGG 1340
DB 51455 MSANGMYKGSKWWYTYWYTGATKKYA-----GRMKRGACGSKTTCASTRTTYM 51509
QY 1341 GTGCTGTGTGTCGCCCATAGCCAGGCAAGTCTGAGGTGAGCGCCGCTCTCGCCTTGT 1400
DB 51510 STSMMSKMWGTSWYSCAYCAKCTCGMCTCCCMRKRKYKSGMTKMCWCMGYSTYSM 51569
QY 1401 GAGCGAGGACGCTGAGCTGAGCTGCCCTCTCGCTCAACACGCCACTGTCTCTGGAGA 1460
DB 51570 RMRMTSKSGMYTRSMRSMSTKWKSYAMTGMKYTYRRMMWMMWMMWKTYTACKYWC 51629
QY 1461 TCGGGCTGTGCGTGAACATCGGAGCCTGCGCTCGTACGTCCTCGGCAACCCCGACTCCT 1520
DB 51630 WYMSWNRWATRRATWKGKAGYKRWGGTGTWTTMYAYCWCYMWANGSSCMACCCYYC 51689
QY 1521 TCTTCCACGCTGCTACAGC-CTCTGCACCGAGAGTGAGCAGCAGGAGTCTCTCGGTAA 1579
DB 51690 ASMTGGYMSRAKSKWMMCCYCCCMSSMMCYCYCMGSGYRKYTYTSYRACMSCKSAY 51749
QY 1580 TGAGGAGCGGACAGGGCCCGCTGCTGTCAGCAGAGAGAGCTTCTAGAAAGATCCCCCT 1639
DB 51750 SCMYMSKACKSKYCAKSGCYKYYKSMYSTYSRSM-WTSTYSYCMYMSMMWMSYMSW 51808
QY 1640 GGTGCTGAGACAGACTGATGGGCGAGGCTCTGAGGATAGAGGACCGGGGAGAGCCCTCCC 1699
DB 51809 YSKSCYTSYCMRKMWCWCSMYRSARCTWRSWTSCYWSRRSTGWRSKMGWCYRRS 51868

Matches	116;	Conservative	445;	Mismatches	410;	Indels	11;	Gaps	5;																																																																																																																														
Qy	443	CCCTGTGCCCTTTGCTCTGTGCTCTGGGTGTAATCCGGGGCCCAAGAGCTCCCTCTCA	502	Db	17864	CMKAGKWCWYTCGSKKGGKYAYSKRKKGTYYWKRKMSAWSRKKWRWYASMRACGMM	17805	Qy	503	GGGACAGAGGGCAAAGATGGGGCTCACTGGGGCTGCCACCGGGCTTTGGGTGCTGAGG	562	Db	17804	GAACGASRSMKCCWGGYRKSAGSKSRGGTGYMKKGGGKSMSSKKWKGSSSTRRRGS	17745	Qy	563	GGGCTGTGGAGCCCAAGGAGTGGCGTCCCGCCAGCACCGGCACCTCCCGCCC	622	Db	17744	SAKSCSYVMGSMCMSCMSWAKSYVMYCYCYMYRMSYMSYYKYCYSCMGMSSTSYISCC	17685	Qy	623	TCCCCGGCCTCAGCGCTGCACAAAGCAGCT-----TGCCACACCTTCCCAACAAGCC	678	Db	17684	WKMSWGVCYKMKYSGWSSYSTSMGYSSSTCKKYKCSWSSMYKCKTSKYRKRSY	17625	Qy	679	CAGGGCTGAGGCTGCCCAGGACGAGGCTGGGGACCTCTCAGGAGGGGTCCCGG	738	Db	17624	YWGGRKAKKYYCAGRRRMSYWKCAKWNMSYCWMSYCMYYKYSGCTYKSSCTCYKRG	17565	Qy	739	AAGAGGGGCTTCCTTGCCA--CAGAGTCCCTCTCAGCAGGCAGGTGGCGTGCCTCAG	796	Db	17564	GYWGSKTCYSAGGKSRSMYCWMRSSKSSSSMSMARSSWCMGWAGYRKRSGAGWA	17505	Qy	797	CACAGTGTGGGCGGAGGTGCAGGACAAGGTTCCCTTCGCGACCTTAATACCCAGGGTCA	856	Db	17504	GMRSSKGRSTGMRACSKTSYSGTGRSMKKKGYSKYSRMGKGGTCTCYCMKYYKY	17445	Qy	857	GGCAGCGCCAGCGCTGTAGTGGCAACATGGCCCCCTTCAAGAGACCCGCTGTGCAGAGCC	916	Db	17444	RKTSMCWYYMKSWSGKY--RYRCCKKKKGTGTRGMSSKKSGYKSRGMRGSSYSTSC	17388	Qy	917	AGTCACGCTGCGCTGTGTTTTCTGGGGCGCAGCCACTTTTTCTTAACAGGGTGACAAA	976	Db	17387	WKS CWGYSWNKCMKYSYKKRRSRMRGSSKSGWRGYAGRGYCSSSMWSTRKRSSKCSY	17328	Qy	977	CCAGACATCGGGTGGGGGACCTTCAGATATGCCTCTCTAATGGCGGTGCTACTCGC	1036	Db	17327	KSYKGRGKKMGWGMKRGSKYWSMMKKMRSSWSKCYSTKSYSGRRKSGWGRSTKSK	17268	Qy	1037	GTGTGTGGGCAGTCTGGGCATGTGGCTGCGCTTCGGGCGCTGCAGCGCTCCTCCCTC	1096	Db	17267	AKSSMRWAGSKCTYGSSYWSNNRRNRNMKTGCNYMYRRS--RMMXNGNAAAGCTTCCCC	17209	Qy	1097	TCCTGGGGTCTGAGTGTGGGACACCTCTGGCGCGTTTGGGGGCTCGAGAGGAGCCCCA	1156	Db	17208	ANTNNGGGGAAAAGCGCGSASRASCTYKGRMSKSCYRSGTRRRRCWKSKRGRSGRKS	17149	Qy	1157	GTGCCCCACCGCTCGTGGCTCCTCAGCTGCAGCTCCCACTTCTCTGGGGCTGCCAGC	1216	Db	17148	MTRKRSGGKTSYSAKSGRGCTYCWGWRKGRKMRSSRRMKRMYKSYRRKRWMTCKMYC	17089	Qy	1217	CTTCCACTGCCTCAAGGATGCGTCTGAGCCTGGCCTGCACATCCATGCTCCACCGGCTG	1276	Db	17088	SMYMYCRSMCMCKKCCGCGSYWGMSSYSYSGKYSGMKSYWYRSYYSKRSTSKAWR	17029	Qy	1277	GAGGGGTGTTCTGGGGCCTTGGGGACCGTATATCTGGGGAGGGGAGGAGATCGGCCT	1336	Db	17028	SSKRGMTGGRYKGGGRSYGKGGSYKGGGSGWGGK--GGKSTRGSGAGKKSKYKMSCCAR	16970	Qy	1337	AGGGGTGCTGTGTGTGCCCATAGCCAGGCAAGTCTGAGGTGGACGCCCGTCTCTCGCC	1396	Db	16969	YKMSYSKYCYSRWMMCMSSYYCWGGKCTCWSWSSMRKSKGGSWAKGWVWDXGSVTDK	16910	Qy	1397	TTGTGAGCAGGAGCAGCTGAG	1418	Db	16909	SDKMBRSBSKVKSKWMSRRG	16888

RESULT 8
AC018829

LOCUS	AC018829	197360 bp	DNA	linear	HTG 03-FEB-2000			
DEFINITION	Homo sapiens chromosome 3p clone RP11-402P11, WORKING DRAFT SEQUENCE, 33 unordered pieces.							
ACCESSION	AC018829							
VERSION	AC018829.3	GI:6862654						
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.							
SOURCE	Homo sapiens							
ORGANISM	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
REFERENCE	1 (bases 1 to 197360)							
AUTHORS	Li,L., Hu,S., Dong,W., Zhang,X., Wang,J., Zhang,Y., Zhang,H., Liu,B., Bao,W., Liu,B., Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y., Qi,X., Niu,Y., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y., Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D., Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Wang,X., Feng,X., Yu,J. and Yang,H.							
TITLE	Chromosome 3p genomic sequence							
JOURNAL	Unpublished							
REFERENCE	2 (bases 1 to 197360)							
AUTHORS	Li,L., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B., Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y., Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y., Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D., Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Wang,X., Feng,X., Yu,J. and Yang,H.							
TITLE	Direct Submission							
JOURNAL	Submitted (21-DEC-1999) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China							
COMMENT	On Feb 3, 2000 this sequence version replaced gi:6684183. * NOTE: This is a 'working draft' sequence. It currently * consists of 33 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 2243: contig of 2243 bp in length * gap of unknown length * 2244 4381: contig of 2138 bp in length * gap of unknown length * 4382 6628: contig of 2247 bp in length * gap of unknown length * 6629 8467: contig of 1839 bp in length * gap of unknown length * 8468 10758: contig of 2291 bp in length * gap of unknown length * 10759 13438: contig of 2680 bp in length * gap of unknown length * 13439 16301: contig of 2863 bp in length * gap of unknown length * 16302 19585: contig of 3284 bp in length * gap of unknown length * 19586 23956: contig of 4371 bp in length * gap of unknown length * 23957 28296: contig of 4340 bp in length * gap of unknown length * 28297 31554: contig of 3258 bp in length * gap of unknown length * 31555 35178: contig of 3624 bp in length * gap of unknown length * 35179 39632: contig of 4454 bp in length * gap of unknown length * 39633 44084: contig of 4452 bp in length * gap of unknown length * 44085 48826: contig of 4742 bp in length * gap of unknown length * 48827 53193: contig of 4367 bp in length * gap of unknown length * 53194 56825: contig of 3632 bp in length * gap of unknown length							


```

FEATURES
  source
    Location/Qualifiers
      1..145435
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="3p"
        /clone="RP11-439f4"

ORIGIN
Query Match      3.4%; Score 76.4; DB 2; Length 145435;
Best Local Similarity 69.3%; Pred. No. 4.3e-05;
Matches 104; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 119 CTTTTCCTGCGACCTAGTGGTTGTAATAATGAGTCCCTACTCTCTCGGACCATC 178
      |||
Db 79797 CTCCTGTCTCCCACTGCCGATGTCAGATGAGTCCCTACTCTCTCGGACCATC 17856

Qy 179 AAGAGGACATCATGTGATATCAGAGCTGACCTACGACCATGCTGAACCACTACGCA 238
      |||
Db 79857 CGCGGGATGCTCGACTGTGCTCTCTCCAGGATCAGATGATTAACCACTACGCC 79916

Qy 239 AAGACAGCTCTTCCACCAAGGTGAGC 268
      |||
Db 79917 CGGCTGCTCTCTTACCACCAAGGTGGC 79946

RESULT 10
AC022382/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-266J6 map 3p, complete
sequence.
ACCESSION AC022382
VERSION AC022382.4 GI:24942870
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 185067)
  Bao, W., Bao, J., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
  Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z.,
  He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F.,
  Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y.,
  Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S.,
  Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J.,
  Wang, J., Wang, J., Wang, L., Wang, R., Wang, X., Wang, X., Wang, X.,
  Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X.,
  Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,
  Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
  Yu, J. and Yang, H.
  Chromosome 3p genomic sequence
  Unpublished
REFERENCE 2 (bases 1 to 185067)
  Wu, D., Hu, S., Dong, W., Zhang, X., Wang, J., Zhang, Y., Zhang, H.,
  Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y.,
  Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wang, X., Yu, B., Fan, H.,
  Liu, Y., Li, G., Li, C., Bao, J., Bao, J., Wang, X., Song, L., Zhang, L.,
  Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L.,
  Feng, X., Yu, J. and Yang, H.
  Direct Submission
  Submitted (03-FEB-2000) Human Genomic Center, Institute of
  Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
  100101, P.R.China
REFERENCE 3 (bases 1 to 185067)
  Bao, W., Bao, J., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
  Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., He, L.,
  Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, J.,
  Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W.,
  Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M.,
  Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, L.,
  Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F.,
  Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G.,
  Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, Y., Zhang, Y.,
  Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
  Yu, J. and Yang, H.
  Direct Submission
  Submitted (13-NOV-2002) Human Genomic Center, Institute of
  Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
  100101, P.R.China
On Nov 13, 2002 this sequence version replaced gi:12007691.
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgci.gtp.ac.cn
http://www.genomics.org.cn
Contact:hgci@gtp.ac.cn
-----Project Information
Center project name:1k project
Center Clone name: RP11-266J6
-----Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 55% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 39 bases at least Q40
Consensus quality: 184 bases at least Q30
Consensus quality: 447 bases at least Q20
Insert size: 919; sum-of-contigs
Quality coverage: 0.71x in Q20 bases;sum-of-contigs
-----Location/Qualifiers
1..185067
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-266J6"

ORIGIN
Query Match      3.4%; Score 76.4; DB 9; Length 185067;
Best Local Similarity 69.3%; Pred. No. 4.3e-05;
Matches 104; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 119 CTTTTCCTGCGACCTAGTGGTTGTAATAATGAGTCCCTACTCTCTCGGACCATC 178
      |||
Db 19471 CTCCTGTCTCCCACTGCCGATGTCAGATGAGTCCCTACTCTCTCGGACCATC 19412

Qy 179 AAGAGGACATCATGTGATATCAGAGCTGACCTACGACCATGCTGAACCACTACGCA 238
      |||
Db 19411 CGCGGGATGCTCGACTGTGCTCTCTCCAGGATCAGATGATTAACCACTACGCC 19352

Qy 239 AAGACAGCTCTTCCACCAAGGTGAGC 268
      |||
Db 19351 CGGCTGCTCTTACCACCAAGGTGGC 19322

RESULT 11
AC011610/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-266J6, WORKING DRAFT SEQUENCE,
  linear HTG 01-NOV-2000
```


15 unordered pieces.
AC011610
VERSION AC011610.11 GI:11055732
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 189430)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C., Alsebrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.K., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dunn-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulesed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Morgan,M., Morris,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwono,G., Oragune,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Stasone,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,A., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189430)
AUTHORS Worley,K.C.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189430)
AUTHORS Worley,K.C.

COMMENT
Submitted (08-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 31, 2000 this sequence version replaced gi:9719552.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMNT
Center clone name: RP11-266J6
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye 4 of reads
Assembly: Dye-terminator Big Dye 96% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 159633 bases at least Q40
Consensus quality: 177583 bases at least Q30
Consensus quality: 184181 bases at least Q20

Estimated insert size: 183186; sum-of-contigs estimation
Estimated insert size: 198859; agarose-fp estimation
Quality coverage: 3.8x in Q20 bases; agarose-fp estimation
Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 40555: contig of 40555 bp in length
* 40556 40555: gap of unknown length
* 40556 83543: contig of 42888 bp in length
* 83544 83643: gap of unknown length
* 83644 99685: contig of 16042 bp in length
* 99686 99785: gap of unknown length
* 99786 119198: contig of 19413 bp in length
* 119199 119298: gap of unknown length
* 119299 134089: contig of 14791 bp in length
* 134090 134189: gap of unknown length
* 134190 149213: contig of 15024 bp in length
* 149214 149313: gap of unknown length
* 149314 159552: contig of 10239 bp in length
* 159553 159652: gap of unknown length
* 159653 168334: contig of 8682 bp in length
* 168335 168434: gap of unknown length
* 168435 175951: contig of 7517 bp in length
* 175952 176051: gap of unknown length
* 176052 179878: contig of 3827 bp in length
* 179879 179979: gap of unknown length
* 179980 184704: contig of 4726 bp in length
* 184705 184804: gap of unknown length
* 184805 185811: contig of 1007 bp in length
* 185812 185911: gap of unknown length
* 185912 186919: contig of 1008 bp in length
* 186920 187019: gap of unknown length
* 187020 188304: contig of 1285 bp in length
* 188305 188404: gap of unknown length
* 188405 189430: contig of 1026 bp in length.
* 189431 189430: Location/Qualifiers
* 1.189430
* /organism="Homo sapiens"
* /mol_type="genomic DNA"
* /db_xref="taxon:9606"
* /chromosome="3"
* /clone="RP11-266J6"

ORIGIN
Query Match 3.4%; Score 76.4; DB 2; Length 189430;
Best Local Similarity 69.3%; Pred. No. 4.3e-05;
Matches 104; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Qy 119 CTTTCTTTCTTGGCCAGTCTAGTTGGTAAATAATGAGATGCGGTACCTCTCTGGACCATC 178
Db 120335 CTCTGTCTCTCCCGAGTCCCGCATGGTCCAGATGAGATCCCTTCTTCTGACCACT 120276
Qy 179 AAGAGGGACATCATTTGACTATATCACAGCTGACCTACGACCATGCTGAAACCACTACGCA 238
Db 120275 CGCGGGATGTGCTCGACTGTGCTTCTCTCCAGGATCAGATGATAAACCACTACGCC 120216
Qy 239 AAGACAGCTCTCTTACACCAAGGTGAGC 268
Db 120215 CGGGCTGGTCTCTTTACCAAAAGGTGGC 120186

RESULT 12
AC026196 AC026196 191834 bp DNA linear HTG 29-MAY-2000
LOCUS

DEFINITION Homo sapiens chromosome 3 clone RP11-474F16 map 3p, WORKING DRAFT
SEQUENCE, 22 unordered pieces.

ACCESSION AC026196
VERSION AC026196.3 GI:8101230
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 191834)
Cai, T., Dong, X., Gao, Q., Gu, J., Gong, J., He, F., Kang, N., Lu, T., Ma, Q., Rong, L., Shen, Y., Tan, X., Wang, H., Xi, Y., Xu, Y., Yao, Z., Zheng, Z., Zhu, N., Zhou, X., Zhou, Y., Qiang, B., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, L., Zhang, L., Zhang, M., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Yu, J. and Yang, H.

TITLE Chromosome 3p genomic sequence
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 191834)
Wang, R., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J. and Yang, H.

TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China

COMMENT On May 29, 2000 this sequence version replaced gi:8050894.
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgc.igtp.ac.cn
http://www.genomics.org.cn
Contact:hgc@igtp.ac.cn
----- Project Information
Center project name:1# project
Center clone name: RP11-474F16
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 55% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 185328 bases at least Q40
Consensus quality: 192037 bases at least Q30
Consensus quality: 196045 bases at least Q20
Insert size: 179116; sum-of-contigs
Quality coverage: 5.75x in Q20 bases;sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1632: contig of 1632 bp in length
* 1633 1732: gap of unknown length
* 1733 2953: contig of 1221 bp in length
* 2954 3053: gap of unknown length
* 3054 4934: contig of 1881 bp in length
* 4935 5034: gap of unknown length
* 5035 7618: contig of 2584 bp in length
* 7619 7718: gap of unknown length


```
Query Match          3.4%; Score 76.4; DB 2; Length 222994;
Best Local Similarity 69.3%; Pred. No. 4.2e-05;
Matches 104; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 119 CTTTTCCTTCCAGTGTAGTGGTAAATAAGATGCGGTACCTCTCTGACCAATC 178
Db 13233 CTCCTGTCCTCCAGTCCGATGGTCCAGATGAGATCCCTACTTCTCATCTGGACCACT 13292

Qy 179 AAGAGGACATCATTTGACTATCATCAGCTGACCTACCAAGATGCTGAACCACTAGCA 238
Db 13293 CGGCGGATGTGCTCGACTGCTGCTCTCTCAAGGATCAGATGATAACCACTAGGCC 13352

Qy 239 AAGACAGCCTCTTCCACACCAAGGTGAGC 268
Db 13353 CGGCTGGCTCTTACACACAAGGTGGC 13382

RESULT 15
AC068315 155313 bp DNA linear HTG 27-SEP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-224P21 map 3p, WORKING DRAFT
DEFINITION AC068315
AC068315.4 GI:10312236
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155313)
AUTHORS Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,
Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L.,
Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y.,
Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y.,
Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, R.,
Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,
Yang, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H.,
Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,
Zhang, Z., Zhu, B., Yu, J., and Yang, H.
Chromosome 3p genomic sequence
Unpublished
TITLE Chromosome 3p genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155313)
AUTHORS Bao, W., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,
Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,
Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,
Li, G., Li, C., Bao, Q., Bao, Q., Bao, J., Song, L., Zhang, L., Guo, D.,
Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.,
and Yang, H.
Direct Submission
JOURNAL Submitted (02-MAY-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On Sep 27, 2000 this sequence version replaced gi:8101156.
COMMENT -----Genome Center
Center:Beijing Center
Center code:Beijing
website:http://hg.cigtp.ac.cn
http://www.genomics.org.cn
Contact:hgc@igtp.ac.cn
----- Project Information
Center project name:1k project
Center Clone name: RP11-224P21
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 55% of reads
Assembly program: Dye-terminator Big Dye; 45% of reads
Consensus quality: 154843 bases at least Q40
Consensus quality: 165011 bases at least Q30
Consensus quality: 168985 bases at least Q20
Insert size: 139544; sum-of-contigs
```

Quality coverage: 5.36x in Q20 bases;sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 1139: contig of 1139 bp in length
1140: gap of unknown length
1240: contig of 1209 bp in length
2448: gap of unknown length
2549: contig of 1387 bp in length
3935: contig of 1387 bp in length
4035: gap of unknown length
4336: contig of 2096 bp in length
6131: gap of unknown length
6231: gap of unknown length
7429: contig of 1198 bp in length
7529: gap of unknown length
7430: contig of 1525 bp in length
9055: gap of unknown length
9154: gap of 1796 bp in length
10950: contig of 1796 bp in length
10951: gap of unknown length
11050: gap of unknown length
12637: contig of 1587 bp in length
12638: gap of unknown length
12737: gap of unknown length
15121: contig of 2384 bp in length
15122: gap of unknown length
15221: gap of unknown length
15222: contig of 1999 bp in length
17220: contig of 1999 bp in length
17221: gap of unknown length
17320: gap of unknown length
20436: contig of 3116 bp in length
20437: gap of unknown length
20536: gap of unknown length
23761: contig of 3225 bp in length
23762: gap of unknown length
26190: contig of 2329 bp in length
26290: gap of unknown length
28871: contig of 2581 bp in length
28971: gap of unknown length
33062: contig of 4091 bp in length
33162: gap of unknown length
37033: contig of 3871 bp in length
37133: gap of unknown length
43465: contig of 6332 bp in length
43565: gap of unknown length
48805: contig of 5240 bp in length
48905: gap of unknown length
56158: contig of 7253 bp in length
56258: gap of unknown length
63409: contig of 7151 bp in length
63509: gap of unknown length
72539: contig of 9030 bp in length
72639: gap of unknown length
83389: contig of 10750 bp in length
83489: gap of unknown length
91426: contig of 7937 bp in length
91526: gap of unknown length
99509: contig of 7983 bp in length
99609: gap of unknown length
108238: contig of 8629 bp in length
108338: gap of unknown length
119079: contig of 10741 bp in length
119179: gap of unknown length
135925: contig of 16746 bp in length
135926: gap of unknown length
136025: gap of unknown length
155313: contig of 19288 bp in length.
```

FEATURES

source

```
1. 155313
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-224P21"
```



```
misc_feature 1. .1139
misc_feature /note="assembly_name:Contig32"
misc_feature 1240. .2448
misc_feature /note="assembly_name:Contig33"
misc_feature 2549. .3935
misc_feature /note="assembly_name:Contig34"
misc_feature 4036. .6131
misc_feature /note="assembly_name:Contig35"
misc_feature 6232. .7429
misc_feature /note="assembly_name:Contig36"
misc_feature 7530. .9054
misc_feature /note="assembly_name:Contig37"
misc_feature 9155. .10950
misc_feature /note="assembly_name:Contig38"
misc_feature 11051. .12837
misc_feature /note="assembly_name:Contig39"
misc_feature 12738. .15121
misc_feature /note="assembly_name:Contig40"
misc_feature 15222. .17220
misc_feature /note="assembly_name:Contig41"
misc_feature 17321. .20436
misc_feature /note="assembly_name:Contig42"
misc_feature 20537. .23761
misc_feature /note="assembly_name:Contig43"
misc_feature 23862. .26190
misc_feature /note="assembly_name:Contig44"
misc_feature 26291. .28871
misc_feature /note="assembly_name:Contig45"
misc_feature 28972. .33062
misc_feature /note="assembly_name:Contig46"
misc_feature 33163. .37033
misc_feature /note="assembly_name:Contig47"
misc_feature 37134. .43465
misc_feature /note="assembly_name:Contig48"
misc_feature 43566. .48805
misc_feature /note="assembly_name:Contig49"
misc_feature 48906. .56158
misc_feature /note="assembly_name:Contig50"
misc_feature 56259. .63409
misc_feature /note="assembly_name:Contig51"
misc_feature 63510. .72539
misc_feature /note="assembly_name:Contig52
misc_feature clone_end:SP6
misc_feature vector_side:right"
misc_feature 72640. .83389
misc_feature /note="assembly_name:Contig53"
misc_feature 83490. .91426
misc_feature /note="assembly_name:Contig54"
misc_feature 91527. .99509
misc_feature /note="assembly_name:Contig55"
misc_feature 99610. .108238
misc_feature /note="assembly_name:Contig56"
misc_feature 108339. .119079
misc_feature /note="assembly_name:Contig57"
misc_feature 119180. .135925
misc_feature /note="assembly_name:Contig58"
misc_feature 136026. .155313
misc_feature /note="assembly_name:Contig59"
```

ORIGIN

```
Query Match 3.3%; Score 73.2; DB 2; Length 155313;
Best Local Similarity 68.0%; Pred.No. 0.00019;
Matches 102; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 119 CTTTTTCTTGCCAGCTAGGTTGGTAAATAATGAGATGCCGTACTCTCTGGACCATC 178
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
48631 CTCCTGTCTCCCGAGTCCCGATGTCAGATGATGCCCTACTTCTATCTGGACCACT 48690

Qy 179 AAGAGGGACATCATTGACTATACAGCCTGACCTACGACGATGCTGAACCACTACGCA 238
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
48691 CGCGGGATGTGTCGACTGAGCTTCTCTCCAAGGATGATGATTAACCACTACCCC 48750

Qy 239 AAGACAGCCTCTTCCACCACCAAGGTGAGC 268
```

```
Db ... 48751 CGGGCTGGCTCTCTTTACCAACAAGGTGGGC 48780

Search completed: September 25, 2005, 11:27:01
Job time : 10407.7 secs
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 01:51:04 ; Search time 1359.73 Seconds
(without alignment)
9756.464 Million cell updates/sec

Title: US-10-615-659-27

Perfect score: 2241

Sequence: 1 tgcctgacagacccggc.....gcctccacaggggccctccc 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_16Dec04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002as:*

7: geneseqn2002bs:*

8: geneseqn2003as:*

9: geneseqn2003bs:*

10: geneseqn2003cs:*

11: geneseqn2003ds:*

12: geneseqn2004as:*

13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2241	100.0	2241	12	ADJ93378 Human BGS
2	2241	100.0	101270	12	ADQ17814 Human sof
3	1795.8	80.1	3465	12	ADJ93363 Human BGS
4	1742.8	77.8	3554	12	ADJ93364 Human tub
5	272.4	12.2	1838	12	ADJ93357 Human BGS
6	272.4	12.2	1939	12	ADJ93361 Human BGS
7	138.4	6.2	1859	12	ADJ93362 Human BGS
8	71.6	3.2	5282	12	ADQ64426 Novel hum
9	69.2	3.1	452	6	ABQ97805 Mouse ES
10	67.8	3.0	1042	5	AA572428 DNA encod
11	67.8	3.0	2250	5	AA569383 DNA encod
12	60.8	2.7	5452	10	ADC86736 Human GPC
13	58.2	2.6	629	13	ACN54594 Corton an
14	58	2.6	1337	2	AZL17263 Human gen
15	57.4	2.6	588	13	ACN54596 Cotton an
16	57	2.5	1327	6	ABQ68452 Listeria
17	56.8	2.5	3163	10	ADC87060 Human GPC
18	56	2.5	12733	6	ABK98631 Vector pE
19	56	2.5	12733	9	ACD13882 L. lactis
20	56	2.5	12739	6	ABK98592 Vector pE

21	56	2.5	12739	9	ACD13843	Accl13843 Plasmid p
22	55.6	2.5	1000	3	AAA02484	Aaa02484 Human col
23	55.2	2.5	2000	8	ADA71938	Ada71938 Rice gene
24	55	2.5	3133	10	ADC86738	Adc86738 Human GPC
25	55	2.5	28198	10	ADG37080	Adg37080 Mouse pla
26	54.4	2.4	110000	12	ADQ97050_0	Adq97050 Human can
27	54.4	2.4	114955	2	AAX53491	Aax53491 Human ade
28	54.2	2.4	931	11	ACN87435	Acn87435 Breast ca
29	53.8	2.4	600	6	ABQ52497	Abq52497 Oligonuc1
30	53.8	2.4	600	6	ABQ52496	Abq52496 Oligonuc1
31	53.6	2.4	2188	2	AAZ77506	Aaz77506 Human ova
32	53.2	2.4	1065	6	ABT09682	Abt09682 Human pva
33	53.2	2.4	40491	11	ACN44798	Acn44798 Human gen
34	52.8	2.4	922	10	ADC86708	Adc86708 Human GPC
35	52.8	2.4	3198	2	AAx02974	Aax02974 Human lu-
36	52.2	2.3	318	3	AAA38184	Aaa38184 Primer us
37	52.2	2.3	778	4	AAH08565	Aah08565 Human cDN
38	52.2	2.3	2326	4	AAH16735	Aah16735 Human cDN
39	51.2	2.3	320	3	AAA38185	Aaa38185 Primer us
40	51.2	2.3	1117	10	ADC86688	Adc86688 Human GPC
41	51	2.3	2561	5	AAH28500	Aah28500 Rabbit lo
42	50.6	2.3	500	12	ADP85918	Adp85918 Synthetic
43	50.6	2.3	1218	3	AAA02488	Aaa02488 Human col
44	50	2.2	320	3	AAA38186	Aaa38186 Primer us
45	49.8	2.2	420	13	ACN51217	Acn51217 Cotton an

ALIGNMENTS

RESULT 1

ADJ93378

ID ADJ93378 standard; DNA; 2241 BP.

XX

AC ADJ93378;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human BGS-42 gene promoter DNA sequence.

XX

KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nocotropic; antiparkinsonian; antarthritic; antiseborrheic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.

XX Homo sapiens.

XX WO2004005487-A2.

XX

XX 15-JAN-2004.

XX

XX 09-JUL-2003; 2003WO-US021605.

XX

XX 09-JUL-2002; 2002US-0394725P.

XX

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX

XX Feder JN, Wu S, Nelson TC;

XX

XX WPI; 2004-099381/10.

XX

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide, e.g.

XX

XX useful for preventing, treating or ameliorating a medical condition, e.g.

aberrant cellular proliferation, reproductive disorders or testicular disorders.

Claim 1; SEQ ID NO 27; 343pp; English.

This invention relates to a novel testis-specific tubulin tyrosine-ligase-like polypeptide, designated the BGS-42 polypeptide. The invention may be useful for the development of compounds with a cyostatic, respiratory

-Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen, anti-inflammatory, anabolic, hypertensive, osteopathic, nootropic, antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial, immunosuppressive, antiseborrheic or dermatological activity acting as tyrosine ligase modulators. In addition, the disclosed sequences may be used for gene therapy. The BGS-42 polypeptide or polynucleotide can be used for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject, and for preventing, treating or ameliorating a medical condition, such as a disorder related to aberrant tubulin ligase activity, a disorder related to aberrant tubulin-carboxypeptidase activity, aberrant cellular proliferation, reproductive disorders, testicular disorders, testicular cancer, pulmonary disorders, lung cancer, gastrointestinal disorders, colon cancer, stomach cancer, neural disorders, brain cancer, liver cancer, or proliferative condition of the testis, lung, small intestine, brain or lymph tissue. The BGS-42 polypeptide, polynucleotide, or their modulators are also useful for treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS-42 polypeptide can be used as a preventive agent for immunological disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's disease or scleroderma. The antibodies may be used to purify, detect and target the BGS-42 polypeptides. The present sequence is that of a gene which encodes the human BGS-42 protein of the invention. The present sequence is that of the promoter of the gene which encodes the human BGS-42 sequence of the invention.

Sequence 2241 BP; 388 A; 598 C; 769 G; 386 T; 0 U; 0 Other;

Query Match		100.0%;	Score 2241;	DB 12;	Length 2241;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 2241;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TGCGTGAGAGAGCCGGCTGGGATCTCAGCTGACAGCCATTTCTCTGATGGTT	60		
Db	1	TGCGTGAGAGAGCCGGCTGGGATCTCAGCTGACAGCCATTTCTCTGATGGTT	60		
Qy	61	GCCTCACTCCATCCCTGGCTGGTGAACCTTGAATGCTGGGCATCAATAAGACT	120		
Db	61	GCCTCACTCCATCCCTGGCTGGTGAACCTTGAATGCTGGGCATCAATAAGACT	120		
Qy	121	TTTTTCTTGCCAGTCTAGTTGGTAAATAATGAGATGCCGTA	180		
Db	121	TTTTTCTTGCCAGTCTAGTTGGTAAATAATGAGATGCCGTA	180		
Qy	181	GAGGGACATATTACTATACAGCCTGACCTAGACAGATGCTGAACCATACGCAA	240		
Db	181	GAGGGACATATTACTATACAGCCTGACCTAGACAGATGCTGAACCATACGCAA	240		
Qy	241	GACAGCTCTTACACCAAGGTGAGCCGCGCGCTTGAGCGTGGGGAGCCGGCAA	300		
Db	241	GACAGCTCTTACACCAAGGTGAGCCGCGCGCTTGAGCGTGGGGAGCCGGCAA	300		
Qy	301	GGGTGGTGGGAGGAGCTTTCAGCCATTAGGACCCCTCGTGGCTGGTCAAGTGGCCAC	360		
Db	301	GGGTGGTGGGAGGAGCTTTCAGCCATTAGGACCCCTCGTGGCTGGTCAAGTGGCCAC	360		
Qy	361	CAGTCACTCTCTGGTGCATCCAGGACTCTGTGCTTAAGCCGTCGCCAGAACCTCG	420		
Db	361	CAGTCACTCTCTGGTGCATCCAGGACTCTGTGCTTAAGCCGTCGCCAGAACCTCG	420		
Qy	421	GTGCCACACCCACCCAGCACCCCTGTGCCCTTTTGTCTGTGCTCTGGGTGAATCCGG	480		
Db	421	GTGCCACACCCACCCAGCACCCCTGTGCCCTTTTGTCTGTGCTCTGGGTGAATCCGG	480		
Qy	481	GGCCCCAGAGCTCCCTCTCAGGGCACAGAGGCCAAAGATGGGGGCTGACTGGGGGCTGC	540		

Db	481	GGCCCCAGAGCTCCCTCTCAGGGCACAGAGGCCAAAGATGGGGCTGACTGGGGGCTGC	540		
Qy	541	CACCGGGCTTTGGTGTCTGAGGGGGCTGTGGGACCCAGGGGGAAGAGTGCCTCCCTCC	600		
Db	541	CACCGGGCTTTGGTGTCTGAGGGGGCTGTGGGACCCAGGGGGAAGAGTGCCTCCCTCC	600		
Qy	601	CAGCACCGGGCACTCCCCCGCTCCCGGCTCCAGCCCTGCACAAAGCAGCTTGCCAC	660		
Db	601	CAGCACCGGGCACTCCCCCGCTCCCGGCTCCAGCCCTGCACAAAGCAGCTTGCCAC	660		
Qy	661	ACCTTTCCACCAAGGGCTGAGGCTTCCCGAGGAGCGAGGCTGTGGGAGACCTG	720		
Db	661	ACCTTTCCACCAAGGGCTGAGGCTTCCCGAGGAGCGAGGCTGTGGGAGACCTG	720		
Qy	721	CTGAGGAGGGGTCCCGGAAGAGGGCTTCCCTGGCACAGAGTCCCTCTCAGCAGGCCA	780		
Db	721	CTGAGGAGGGGTCCCGGAAGAGGGCTTCCCTGGCACAGAGTCCCTCTCAGCAGGCCA	780		
Qy	781	GGTGGCGCTGCCTCAGCACAGTGTGGGGCGAGGTGCAGCAAGGTTCCTTCCCGCACC	840		
Db	781	GGTGGCGCTGCCTCAGCACAGTGTGGGGCGAGGTGCAGCAAGGTTCCTTCCCGCACC	840		
Qy	841	TAATACCCAGGGTCAAGCCAGCCAGCCAGCCCTGCTAGTAGTGGCAACATGGCCCTTCAAGA	900		
Db	841	TAATACCCAGGGTCAAGCCAGCCAGCCAGCCCTGCTAGTAGTGGCAACATGGCCCTTCAAGA	900		
Qy	901	CCCGGTGTGAGAGCCAGTCAAGCGTGGTGTGTTCCTGGGGCCAGCCACTTTTTCCTCC	960		
Db	901	CCCGGTGTGAGAGCCAGTCAAGCGTGGTGTGTTCCTGGGGCCAGCCACTTTTTCCTCC	960		
Qy	961	TTAAACAGGGTGAACAAACAGACATCGGGGTGCGGGGACTTCAAGATATGCTCTTAATG	1020		
Db	961	TTAAACAGGGTGAACAAACAGACATCGGGGTGCGGGGACTTCAAGATATGCTCTTAATG	1020		
Qy	1021	GCGGGTGCTACCTCGCGTGTGGTGGGAGTGTGGGAGTGTGGGATGTGGTCCGGGCCCTG	1080		
Db	1021	GCGGGTGCTACCTCGCGTGTGGTGGGAGTGTGGGAGTGTGGGATGTGGTCCGGGCCCTG	1080		
Qy	1081	CAGCGCT	1140		
Db	1081	CAGCGCT	1140		
Qy	1141	CCTGGAGGAGCCCAAGTGCACCCAGCCCTCGGTGGCTCAGCTGACGCTCCCAATTC	1200		
Db	1141	CCTGGAGGAGCCCAAGTGCACCCAGCCCTCGGTGGCTCAGCTGACGCTCCCAATTC	1200		
Qy	1201	CTGGGGCTGCCAGCCCTCCT	1260		
Db	1201	CTGGGGCTGCCAGCCCTCCT	1260		
Qy	1261	ATGCTCCACCGGTGGAGGGGTGTCTCTGGGGCCCTGGGGGACCGGTATATCTGGGGAGG	1320		
Db	1261	ATGCTCCACCGGTGGAGGGGTGTCTCTGGGGCCCTGGGGGACCGGTATATCTGGGGAGG	1320		
Qy	1321	GGGAGGAGTCCGCTTAGGGGTGCTGTGGTGTCCCATAGCCAGGCAAGTGTGAGGTGG	1380		
Db	1321	GGGAGGAGTCCGCTTAGGGGTGCTGTGGTGTCCCATAGCCAGGCAAGTGTGAGGTGG	1380		
Qy	1381	ACGGCCGCTCTCGCTTGTGAGGAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG	1440		
Db	1381	ACGGCCGCTCTCGCTTGTGAGGAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG	1440		
Qy	1441	CAGCCACTGTCTCTGGAGATCGGGTGTGCGTGAAACATCGGAGCTGCGCTGCTGACGT	1500		
Db	1441	CAGCCACTGTCTCTGGAGATCGGGTGTGCGTGAAACATCGGAGCTGCGCTGCTGACGT	1500		
Qy	1501	CCCGGCCAACCCGACTCTTCTTCCCA	1560		
Db	1501	CCCGGCCAACCCGACTCTTCTTCCCA	1560		
Qy	1561	GCAGGATTCCTGGGTAAAGTGAGAGACGAGGGGCCCGGCTGCTGCTGAGAGAGG	1620		

QY 661 ACCCTTCCCAACAGGCCAGGGCTGAGGCCCTGCCAGGACGACAGGCTGTGGGACCCCTG 720
DB 59478 ACCCTTCCCAACAGGCCAGGGCTGAGGCCCTGCCAGGACGACAGGCTGTGGGACCCCTG 59419
QY 721 CTGAGGAGGGGTCCGGAAGAGGGGCTTCCCTGGCAAGAGGTCCTCTCAGCAGGCCA 780
DB 59418 CTGAGGAGGGGTCCGGAAGAGGGGCTTCCCTGGCAAGAGGTCCTCTCAGCAGGCCA 59359
QY 781 GGTCCGGCTGCTCAGCACAGTGTGGGCGGAGGTGCAGGACAGGTCCTCTCCGCAAC 840
DB 59358 GGTCCGGCTGCTCAGCACAGTGTGGGCGGAGGTGCAGGACAGGTCCTCTCCGCAAC 59299
QY 841 TAATACCCCAAGGTCAGGCCAGCCAGCGCTGCTAGTGGCAACATGGCCCTCTCAAAGA 900
DB 59298 TAATACCCCAAGGTCAGGCCAGCCAGCGCTGCTAGTGGCAACATGGCCCTCTCAAAGA 59239
QY 901 CCCCGTGTGAGAGCCAGTCAGCGTCGCTGTGTTTTCTGGGGGCCAGCACTTTTTC 960
DB 59238 CCCCGTGTGAGAGCCAGTCAGCGTCGCTGTGTTTTCTGGGGGCCAGCACTTTTTC 59179
QY 961 TTAACAGGTCACAAACAGACATCGGGGTGCGGGACTTTCAGATATGCTCTCTAATG 1020
DB 59178 TTAACAGGTCACAAACAGACATCGGGGTGCGGGACTTTCAGATATGCTCTCTAATG 59119
QY 1021 GCGCGTGTACCTCGCGTGTGGGCACTGCTGGGCATGTGGCTGCTCCGGGCCCTG 1080
DB 59118 GCGCGTGTACCTCGCGTGTGGGCACTGCTGGGCATGTGGCTGCTCCGGGCCCTG 59059
QY 1081 CAGCGCTCTCTCTCTCTGGGGTCTGAGGTGTGGACACCTGTGGCCCGTTTGGGG 1140
DB 59058 CAGCGCTCTCTCTCTCTGGGGTCTGAGGTGTGGACACCTGTGGCCCGTTTGGGG 58999
QY 1141 CCTGGAGGAGCCCACTGACCCAGCCCGCTCGGTGCGCTCAGCTGCAGCCTCCCATTC 1200
DB 58998 CCTGGAGGAGCCCACTGACCCAGCCCGCTCGGTGCGCTCAGCTGCAGCCTCCCATTC 58939
QY 1201 CTGGGGCTGTCCAGCCCTCTCACTGCTCAAGATGCGTCTGAGCCTTGGCTTGCACATCC 1260
DB 58938 CTGGGGCTGTCCAGCCCTCTCACTGCTCAAGATGCGTCTGAGCCTTGGCTTGCACATCC 58879
QY 1261 ATGCTCCACCGGTGAGGGGTGTCTGGGGCCCTGGGGGACCGTATATCTGGGGAGG 1320
DB 58878 ATGCTCCACCGGTGAGGGGTGTCTGGGGCCCTGGGGGACCGTATATCTGGGGAGG 58819
QY 1321 GGGAGGATCCGCCCTAGGGGTGCTGTGTGTCGCCCATAGCCAGGCAAGTCTGAGTGG 1380
DB 58818 GGGAGGATCCGCCCTAGGGGTGCTGTGTGTCGCCCATAGCCAGGCAAGTCTGAGTGG 58759
QY 1381 ACGCCCGTCTCTCGCTTGTGAGCGAGCAGAGCTGAGCTGAGCTGCCCTCTCTGCCCTAA 1440
DB 58758 ACGCCCGTCTCTCGCTTGTGAGCGAGCAGAGCTGAGCTGAGCTGCCCTCTCTGCCCTAA 58699
QY 1441 CAGCCCACTGTCTTGAGATCGGGCTGTGCGTGAACATGCGGAGCGCTGCCCTGTGACGT 1500
DB 58698 CAGCCCACTGTCTTGAGATCGGGCTGTGCGTGAACATGCGGAGCGCTGCCCTGTGACGT 58639
QY 1501 CCGGGCAACCCGACTCTCTTCCACCGCTGTACAGCCTCTGCACCGAGAGTGAGCA 1560
DB 58638 CCGGGCAACCCGACTCTCTTCCACCGCTGTACAGCCTCTGCACCGAGAGTGAGCA 58579
QY 1561 GCAGGAGTCTCTGGGTAACTGAGGAGACGGCAGAGGGCCCCAGTGTCTCAGCAGAGAGG 1620
DB 58578 GCAGGAGTCTCTGGGTAACTGAGGAGACGGCAGAGGGCCCCAGTGTCTCAGCAGAGAGG 58519
QY 1621 CTTCTAGAAATCCCCCTGTGTGAGACAGACTGTATGGGGCAGGGTCTGAGGATAGAG 1680
DB 58518 CTTCTAGAAATCCCCCTGTGTGAGACAGACTGTATGGGGCAGGGTCTGAGGATAGAG 58459
QY 1681 GACCGGGAGAGGCTCCCATGTCATGTCATGCGAGTACAGAGGCCAGGGGCCCGGG 1740
DB 58458 GACCGGGAGAGGCTCCCATGTCATGTCATGCGAGTACAGAGGCCAGGGGCCCGGG 58399

QY 1741 AGGAGGGGGGCGAGTCAAGAGCTGTGTGTTCTGTCTAGGCTTGTGGAACGCGCGCG 1800
DB 58398 AGGAGGGGGGCGAGTCAAGAGCTGTGTGTTCTGTCTAGGCTTGTGGAACGCGCGCG 58339
QY 1801 AGCAGCTGGGGCAAGCGCGCTCCGTGTCGAGGCGGAGGGGTGACGTGGAGCGGCCG 1860
DB 58338 AGCAGCTGGGGCAAGCGCGCTCCGTGTCGAGGCGGAGGGGTGACGTGGAGCGGCCG 58279
QY 1861 AGTCACAGAGACACTGCAGGGGAAGCGCAGGACAGACTGGGGGTGTGGGAGCAGGCTCG 1920
DB 58278 AGTCACAGAGACACTGCAGGGGAAGCGCAGGACAGACTGGGGGTGTGGGAGCAGGCTCG 58219
QY 1921 GCCCATTCGCTATAGGGCGGAAGTGGGGCGGGTTGGGAGCCTCCGTGGCCCTGGCC 1980
DB 58218 GCCCATTCGCTATAGGGCGGAAGTGGGGCGGGTTGGGAGCCTCCGTGGCCCTGGCC 58159
QY 1981 CATCTCTCAGTCCCACTGCTGCTCGGACAGATAGGCGGAGGCTGTGCTGTCTTTCAG 2040
DB 58158 CATCTCTCAGTCCCACTGCTGCTCGGACAGATAGGCGGAGGCTGTGCTGTCTTTCAG 58099
QY 2041 AAGACTTCCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACAGAGCT 2100
DB 58098 AAGACTTCCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGGTCAGCCACAGAGCT 58039
QY 2101 GCAGCAGGACGACGACGAAGCAAGCCAGGACACAGAGGAGGAGCGGAGCAGCGAAC 2160
DB 58038 GCAGCAGGAGCAGCAGGAAGCAAGCCAGGACACAGAGGAGGAGCGGAGCAGCGAAC 57979
QY 2161 TGAGCAGCAGGCAAGGTGCGCTGGGCGGGGAGGACAGTGCAGGCCACAGAGCTCGG 2220
DB 57978 TGAGCAGCAGGCAAGGTGCGCTGGGCGGGGAGGACAGTGCAGGCCACAGAGCTCGG 57919
QY 2221 GGCTCCACAGGGGCGCTCC 2241
DB 57918 GGCTCCACAGGGGCGCTCC 57898

RESULT 3
ADJ93363
ID ADJ93363 standard; DNA; 3465 BP.
AC ADJ93363;
XX
DT 06-MAY-2004 (first entry)
XX Human BGS-42 protein-related DNA clone C SeqID11.
DE
XX testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's disease;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.
OS Homo sapiens.
XX
XX WO2004005487-A2.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;
PI WPI; 2004-099381/10.
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
PT useful for preventing, treating or ameliorating a medical condition, e.g.
PT aberrant cellular proliferation, reproductive disorders or testicular
PT disorders.
XX
XX Example 4; SEQ ID NO 11; 343pp; English.
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -gen, gastrointestinal-gen, neuroprotective, endocrine-gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, neurotropic,
CC antiparkinsonian, antithrombotic, antiaesthetic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.
XX
SQ Sequence 3465 BP; 667 A; 1074 C; 1126 G; 598 T; 0 U; 0 Other;

Query Match
Beat Local Similarity 80.1%; Score 1795.8; DB 12; Length 3465;
Matches 1797; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 381 CCCAGGACTCTCTGCTTAAGGCGCGTGGCCAGAAATCACTCGGTGCCACCCACCCCGCAGC 440
DB 1 CCCAGGACTCTCTGCTTAAGGCGCGTGGCCAGAAATCACTCGGTGCCACCCACCCCGCAGC 60
QY 441 ACCCTGTGCCCTTTGCTCTGTCTGTCTGTGGTGAATTCGGGGGCCCCAGAGTCCCTCT 500
DB 61 ACCCTGTGCCCTTTGCTCTGTCTGTCTGTGGTGAATTCGGGGGCCCCAGAGTCCCTCTCT 120
QY 501 CAGGCGACAGAGCCCAAGATGGGCTGCTACCTGGGGGCTGCCACCGGGCTTTGGGTGCTGA 560
DB 121 CAGGCGACAGAGCCCAAGATGGGCTGCTACCTGGGGGCTGCCACCGGGCTTTGGGTGCTGA 180
QY 561 GGGGGCTGTGGGACCCCGAGGGAAGAGGTGCCGCTCCCCCGACAGCCGGGCACTCCCCCG 620
DB 181 GGGGGCTGTGGGACCCCGAGGGAAGAGGTGCCGCTCCCCCGACAGCCGGGCACTCCCCCG 240
QY 621 CTTCCCGCGCTCCAGCCCTGCAAAAGAGCTTTGCCACACCTTTCCCAACAGAGGCCCA 680
DB 241 CTTCCCGCGCTCCAGCCCTGCAAAAGAGCTTTGCCACACCTTTCCCAACAGAGGCCCA 300
QY 681 GGGCTGAGCCCTGCCCAGAGCCAGGGGTGGGGACCTTCTGAGGGAGGGGTCCCGGAA 740
DB 301 GGGCTGAGCCCTGCCCAGAGCCAGGGGTGGGGACCTTCTGAGGGAGGGGTCCCGGAA 360
QY 741 GAGGGGCTTCCCTGGCAGAGAGGTCCCTCTCAGCAGGCGCAGGTGCGGCTGCTCAGCACA 800
DB 361 GAGGGGCTTCCCTGGCAGAGAGGTCCCTCTCAGCAGGCGCAGGTGCGGCTGCTCAGCACA 420

QY 801 GTGTGGGGCGGAGGTGCAGGACAAGTTTCCCTTCGCAACCTAATAATACCCAGGGTCAGGCC 860
DB 421 GTGTGGGGCGGAGGTGCAGGACAAGTTTCCCTTCGCAACCTAATAATACCCAGGGTCAGGCC 480
QY 861 AGCGCCAGCGCTGTCTAGTGGCAACATATGGCCCTTTCAAAAGACCCCGTGTGCAGAGCCAGTTC 920
DB 481 AGCGCCAGCGCTGTCTAGTGGCAACATATGGCCCTTTCAAAAGACCCCGTGTGCAGAGCCAGTTC 540
QY 921 AGCGTGCCTGTGTCTTCTGGGGCGCAGCCACTTTTCTTTTCTTAAACAGGGTGACAAACAG 980
DB 541 AGCGTGCCTGTGTCTTCTGGGGCGCAGCCACTTTTCTTTTCTTAAACAGGGTGACAAACAG 600
QY 981 ACATCGGGGTGCGGGGACTTCACGATATGCTCTCTAATGCGCGTGTCTACCTCCGCTGT 1040
DB 601 ACATCGGGGTGCGGGGACTTCACGATATGCTCTCTAATGCGCGTGTCTACCTCCGCTGT 660
QY 1041 GGTGGGCAGTCTGCGGGCATGTGGCTGCGTTCGGGCGCTTCGAGCGGCTCTCTCTCTCTCT 1100
DB 661 GGTGGGCAGTCTGCGGGCATGTGGCTGCGTTCGGGCGCTTCGAGCGGCTCTCTCTCTCTCTCT 720
QY 1101 GGGGTCTGAGGTGTGGGACACCTCTGTGGCCCGTTTGGGGGCTTCGAGGAGGAGGAGGAGG 1160
DB 721 GGGGTCTGAGGTGTGGGACACCTCTGTGGCCCGTTTGGGGGCTTCGAGGAGGAGGAGGAGG 780
QY 1161 CCACGCGCTCGGTGCGCTCAGCTGCAGCTCCCATTTCCCTGGGGCGCTGCCACGCCCTC 1220
DB 781 CCACGCGCTCGGTGCGCTCAGCTGCAGCTCCCATTTCCCTGGGGCGCTGCCACGCCCTC 840
QY 1221 CACTGCTCTCAAGGATGCGCTCTGAGCTGGCGCTGCACATTCATTCCTCCACCGGCTGAGG 1280
DB 841 CACTGCTCTCAAGGATGCGCTCTGAGCTGGCGCTGCACATTCATTCCTCCACCGGCTGAGG 900
QY 1281 GGGTGTCTGCGGGCGCTGGGGACCGTATATCTGGGAGGGGAGGAGTTCGCGCTCTAGG 1340
DB 901 GGGTGTCTGCGGGCGCTGGGGACCGTATATCTGGGAGGGGAGGAGTTCGCGCTCTAGG 960
QY 1341 GTGCTGTGTGTGTCCTTACGAGGAGGAGTCTGAGGTGGAGCGGCGCTCTCGGCTGTGT 1400
DB 961 GTGCTGTGTGTGTCCTTACGAGGAGGAGTCTGAGGTGGAGCGGCGCTCTCGGCTGTGT 1020
QY 1401 GAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1460
DB 1021 GAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
QY 1461 TCGGGCTGTGCGTGAAACATGCGAGCGCTGCGCTGTGTACGTCCCGGCGAAACCCGCACTCT 1520
DB 1081 TCGGGCTGTGCGTGAAACATGCGAGCGCTGCGCTGTGTACGTCCCGGCGAAACCCGCACTCT 1140
QY 1521 TCTTCCCAACGCTGTACAGCCTCTGCAACGAGGTGAGCAGCAGGAGTTCCTGGGTAAAGT 1580
DB 1141 TCTTCCCAACGCTGTACAGCCTCTGCAACGAGGTGAGCAGCAGGAGTTCCTGGGTAAAGT 1200
QY 1581 GAGGAGAGCGCAGAGGGGCGCCAGTGTCTGAGCAGAGAGGCTTTAGAAAGATCCCGCTG 1640
DB 1201 GAGGAGAGCGCAGAGGGGCGCCAGTGTCTGAGCAGAGAGGCTTTAGAAAGATCCCGCTG 1260
QY 1641 GTGCTGAGACAGACTGATGGGGCAGGGTCTGAGGATAGAGGACCGGGGAGAGGCTCCCA 1700
DB 1261 GTGCTGAGACAGACTGATGGGGCAGGGTCTGAGGATAGAGGACCGGGGAGAGGCTCCCA 1320
QY 1701 TGCTATGGTCAATGGCAGTACAGAGGCGCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1760
DB 1321 TGCTATGGTCAATGGCAGTACAGAGGCGCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
QY 1761 GAGTGTGTGGTCTGTGTCTAGGCTTGTGGACAGCGCGCGCAGCAGCTGCGGGCAGGCGGG 1820
DB 1381 GAGTGTGTGGTCTGTGTCTAGGCTTGTGGACAGCGCGCGCAGCAGCTGCGGGCAGGCGGG 1440
QY 1821 CTCGCTGTGTGGAGCGCGGAGGGGTGCACTGTGAGCGCGCGCAGTCAAGAGACACTGCAAG 1880
DB 1441 CTCGCTGTGTGGAGCGCGGAGGGGTGCACTGTGAGCGCGCGCAGTCAAGAGACACTGCAAG 1500
QY 1881 GAGAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1940

Db 1501 GAGAAAGGCGAGCAGATGGGGGTTGGGAGCAGGCGCTGGGCCCATCGCGCTATGAGGGC 1560
 Qy 1941 GGGAAAGTGGGCGGGTGGGAGAGCTCGGTGGCCCTGGCCCATCTCCAGTCCCGCCAGTCC 2000
 Db 1561 GGGAAAGTGGGCGGGTGGGAGAGCTCGGTGGCCCTGGCCCATCTCCAGTCCCGCCAGTCC 1620
 Qy 2001 TGGCTCGACAGATAGGCGGAGGTGTGCTGCTCTTTTCAGAAAGACTTCCCGCGCAGCATG 2060
 Db 1621 TGGCTCGACAGATAGGCGGAGGTGTGCTGCTCTTTTCAGAAAGACTTCCCGCGCAGCATG 1680
 Qy 2061 GCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTTCACAGGAGCAGCAGAGC 2120
 Db 1681 GCATCCAGCATCTCAAGTGGTGGTTCAGCCACAGAGCTTCACAGGAGCAGCAGAGC 1740
 Qy 2121 AAGCCCGAGGACAGAGGAGGAGCGCGGAGCAGCAGCTTGACGAGCAGGCAAGGTGC 2179
 Db 1741 AAGCCCGAGGACAGAGGAGGAGCGCGGAGCAGCAGCTTGACGAGCAGGCAAGGTGC 1799
 RESULT 4
 ID ADJ93364 standard; DNA; 3554 BP.
 AC ADJ93364;
 DT 06-MAY-2004 (first entry)
 XX Human tubulin tyrosine ligase protein consensus gene sequence SeqID12.
 DE
 XX testis-specific tubulin tyrosine-ligase-like polypeptide;
 KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
 KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
 KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
 KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
 KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
 KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
 KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
 KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
 KW brain cancer; liver cancer; proliferative condition; testis; lung;
 KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
 KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
 KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
 KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ds.
 XX Homo sapiens.
 OS
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1695..3320
 FT /*tag= a
 FT /product= "Human tubulin tyrosine ligase protein"
 XX
 FN WO2004/005487-A2.
 XX
 XX 15-JAN-2004.
 PD
 XX 09-JUL-2003; 2003WO-US021605.
 PF
 XX 09-JUL-2002; 2002US-0394725P.
 PR
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA
 XX Feder JN, Wu S, Nelson TC;
 PI
 XX WPI; 2004-099381/10.
 DR
 XX P-PSDB; ADJ93365.
 DR
 XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
 PT useful for preventing, treating or ameliorating a medical condition, e.g.
 PT aberrant cellular proliferation, reproductive disorders or testicular
 PT disorders.
 XX

PS Example 4; SEQ ID NO 12; 343pp; English.
 XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
 CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
 CC be useful for the development of compounds with a cytosolic, respiratory
 CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
 CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
 CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
 CC immunosuppressive, antiseborrheic or dermatological activity acting as
 CC tyrosine ligase modulators. In addition, the disclosed sequences may be
 CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
 CC used for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, and for preventing, treating or
 CC ameliorating a medical condition, such as a disorder related to aberrant
 CC tubulin ligase activity, a disorder related to aberrant tubulin-
 CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
 CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
 CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
 CC neural disorders, brain cancer, liver cancer, or proliferative condition
 CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
 CC polypeptide, polynucleotide, or their modulators are also useful for
 CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
 CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
 CC -42 polypeptide can be used as a preventive agent for immunological
 CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
 CC disease or scleroderma. The antibodies may be used to purify, detect and
 CC target the BGS-42 polypeptides. The present sequence is that of the
 CC tubulin tyrosine ligase protein consensus gene sequence which was used in
 CC the exemplification of the invention.
 XX
 SQ Sequence 3554 BP; 692 A; 1090 C; 1157 G; 615 T; 0 U; 0 Other;
 Query Match 77.8%; Score 1742.8; DB 12; Length 3554;
 Best Local Similarity 98.2%; Pred. No. 0;
 Matches 1784; Conservative 0; Mismatches 2; Indels 30; Gaps 1;
 Qy 394 GTCTAAGGCGGTGGCGAGATCACTGGTGCACACCCACCCAGCACCCCTGTGCGCT 453
 Db 1 GTCTAAGGCGGTGGCGAGATCACTGGTGCACACCCACCCAGCACCCCTGTGCGCT 60
 Qy 454 TTGCTCTGTCTCTCTGGTGAATCCGGGGCCCGAGAGCTCCCTCTCAGGCGCACAGG 513
 Db 61 TTGCTCTGTCTCTGGTGAATCCGGGGCCCGAGAGCTCCCTCTCAGGCGCACAGG 120
 Qy 514 CCAAGATGAGGCGGTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGAGGGGCTGTGCGA 573
 Db 121 CCAAGATGAGGCGGTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGAGGGGCTGTGCGA 180
 Qy 574 CCCAGGGGGAAGAGTGCCTCCCTCCCGACAGCGGCACTCCCGCGCTCCCGCGGCT 633
 Db 181 CCCAGGGGGAAGAGTGCCTCCCTCCCGACAGCGGCACTCCCGCGCTCCCGCGGCT 240
 Qy 634 CCAGCCCTGCACAAAGCAGCTTCCACACCTTCCCAAGGCGCCAGGCTGAGGCTG 693
 Db 241 CCAGCCCTGCACAAAGCAGCTTCCACACCTTCCCAAGGCGCCAGGCTGAGGCTG 300
 Qy 694 CCCAGGACGAGGCTGTGGGGACCTCTCTGAGGGAGGGGTCCCGGAAGAGGGGCTTCCCT 753
 Db 301 CCCAGGACGAGGCTGTGGGGACCTCTCTGAGGGAGGGGTCCCGGAAGAGGGGCTTCCCT 360
 Qy 754 GGCACAGAGTCCCTCTCAGCAGGCGCAGGTGCGGCTCCCTCAGCAGTGTGGGCGGAG 813
 Db 361 GGCACAGAGTCCCTCTCAGCAGGCGCAGGTGCGGCTCCCTCAGCAGTGTGGGCGGAG 420
 Qy 814 GTCAGGACAGGTTCCCTTCCGACCTAATACCCAGGGTTCAGGCGCAGCGCGCTG 873
 Db 421 GTCAGGACAGGTTCCCTTCCGACCTAATACCCAGGGTTCAGGCGCAGCGCGCTG 480
 Qy 874 CTAGTGGCAACATGGGCCCTTTCAAAGACCCCGTGTGCAGAGCCAGTGCAGCTGCGCTGTG 933
 Db 481 CTAGTGGCAACATGGGCCCTTTCAAAGACCCCGTGTGCAGAGCCAGTGCAGCTGCGCTGTG 540
 Qy 934 TTTTCTGGGGGCGAGCCACTTTTTTCTTAACAGGGGTGACAAACACACATCGGGGTGCG 993

Db 541 TTTTCTGGGGGCGAGCCATCTTTTCTTAAACGGGGTGACAAACAGACATCGGGTGG 600
Qy 994 GGGACTTCAGATATGCTCTCTAATAGCGCGTGTCTACTCCGCTGTGTGGGAGTGTCT 1053
Db 601 GGGACTTCAGATATGCTCTCTAATAGCGCGTGTCTACTCCGCTGTGTGGGAGTGTCT 660
Qy 1054 GGGCATGTGGCTGTCTCGGGCCCTGACAGCGCTCTCCCTCTCTCTGGGGTCTCAGGTG 1113
Db 661 GGGCATGTGGCTGTCTCGGGCCCTGACAGCGCTCTCCCTCTCTCTGGGGTCTCAGGTG 720
Qy 1114 TGGGACACCTGTGGCCCGTTTGGGGGCTTGAGGGAGGCCAGTCCACCGCCCTCGG 1173
Db 721 TGGGACACCTGTGGCCCGTTTGGGGGCTTGAGGGAGGCCAGTCCACCGCCCTCGG 780
Qy 1174 TGCCCTCAGCTGACGCTCCCAATTCCTTGGGGCTTGCGACCGCCCTCACTGCTCAAGG 1233
Db 781 TGCCCTCAGCTGACGCTCCCAATTCCTTGGGGCTTGCGACCGCCCTCACTGCTCAAGG 840
Qy 1234 ATGCGTCTGAGCTGCGCTGCACATCCATGCTCCCAACCGGCTGAGGGGGTGTCTGGG 1293
Db 841 ATGCGTCTGAGCTGCGCTGCACATCCATGCTCCCAACCGGCTGAGGGGGTGTCTGGG 900
Qy 1294 CCTGGGGGACCGTATATCTGGGGAGGGGAGGATCCGCCCTAGGGGTGCTGTGGTG 1353
Db 901 CCTGGGGGACCGTATATCTGGGGAGGGGAGGATCCGCCCTAGGGGTGCTGTGGTG 960
Qy 1354 CCCATAGCCAGGCAAGTCTGAGGTGAGCGGCCCTCTCTCGCTTGTGAGGAGGACAG 1413
Db 961 CCCATAGCCAGGCAAGTCTGAGGTGAGCGGCCCTCTCTCGCTTGTGAGGAGGACAG 1020
Qy 1414 CTGAGCTGAGCTGCGCTCTGCTGCTTAAACAGCCACTGCTCTGAGAGTCGGGTGCTG 1473
Db 1021 CTGAGCTGAGCTGCGCTCTGCTGCTTAAACAGCCACTGCTCTGAGAGTCGGGTGCTG 1080
Qy 1474 GAACATGGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1533
Db 1081 GAACATGGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Qy 1534 CTACAGCTCTGCAACCGAGAGTGAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1593
Db 1141 CTACAGCTCTGCAACCGAGAGTGAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Qy 1594 AGGGCCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1653
Db 1201 AGGGCCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Qy 1654 CTGATGGGGGAGGCTGCTGAGGATAGAGACCGGGGAGGCTCCCATGCTCATGCTCAT 1713
Db 1261 CTGATGGGGGAGGCTGCTGAGGATAGAGACCGGGGAGGCTCCCATGCTCATGCTCAT 1320
Qy 1714 GGCAGTACAGAGGCGGGGCTCCGGGAGGGGAGGCGGGGAGTCAAGGAGTGTGTGGTTC 1773
Db 1321 GGCAGTACAGAGGCGGGGCTCCGGGAGGGGAGGCGGGGAGTCAAGGAGTGTGTGGTTC 1380
Qy 1774 TGTGCTAGGCTTGTGACAGCGCGCGAGCTGCGGGAGGCGCGGCTCCGTGCTCGGA 1833
Db 1381 TGTGCTAGGCTTGTGACAGCGCGCGAGCTGCGGGAGGCGCGGCTCCGTGCTCGGA 1440
Qy 1834 GGGCCAGGGGTGAGCTGAGCGGGCGAGTCAACAGACACTGACAGGAGAGGCGAGCG 1893
Db 1441 GGGCCAGGGGTGAGCTGAGCGGGCGAGTCAACAGACACTGACAGGAGAGGCGAGCG 1500
Qy 1894 AGACTGGGGGTGTGGGAGCAGCTTGGGCGGCTCGGCTATGAGGCGGGGAGTGGGGCG 1953
Db 1501 AGACTGGGGGTGTGGGAGCAGCTTGGGCGGCTCGGCTATGAGGCGGGGAGTGGGGCG 1560
Qy 1954 GGTGGGAGGCTCGTGGCGCTTGGCGGCTTCAAGTCCCGAGTCTGGCTCGGACAGA 2013
Db 1561 GGTGGGAGGCTCGTGGCGCTTGGCGGCTTCAAGTCCCGAGTCTGGCTCGGACAGA 1620
Qy 2014 TAGGCGAGGCTGCTGCTGCTTCA-----GAAG 2043

Db 1621 TAGGCGAGGCTGTGCTGCTCTTTCAGAGTAGCAGCGTGGGAAGAAGGAGTTCTCTGAAG 1680
Qy 2044 ACTTCCGGCGCACCATGTCATCCAGCATCTCTCAAGTGGTGGTCAAGTGGTGGTCA 2103
Db 1681 ACTTCCGGCGCACCATGTCATCCAGCATCTCTCAAGTGGTGGTCAAGTGGTGGTCA 1740
Qy 2104 GCAGGAGCAGCAGCAAGCAAGCCAGGAGCAGAGGAGGAGGCGCGGAGCAGCAGCTGA 2163
Db 1741 GCAGGAGCAGCAGCAAGCAAGCCAGGAGCAGAGGAGGAGGCGCGGAGCAGCAGCTGA 1800
Qy 2164 GCAGGAGCAGCAAGGTGC 2179
Db 1801 GCAGGAGCAGCAAGTGC 1816
RESULT 5
ADJ93357
ID ADJ93357 standard; cDNA; 1838 BP.
XX
AC ADJ93357;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human BGS-42 cDNA sequence SeqID1.
XX
KW testis-specific tubulin tyrosine-ligase-like polypeptide;
KW BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
KW neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
KW osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
KW anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
KW dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
KW tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
KW testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
KW gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
KW brain cancer; liver cancer; proliferative condition; testis; lung;
KW small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
KW emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
KW Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;
KW sepsis; acne; Sjogren's disease; scleroderma; human; gene; ss.
XX
OS Homo sapiens.
FH
Key Location/Qualifiers
FT CDS 153..1778
FT /*tag= a
FT /product= "Human BGS-42 protein"
XX
XX WO2004005487-A2.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2003; 2003WO-US021605.
XX
XX 09-JUL-2002; 2002US-0394725P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Feder JN, Wu S, Nelson TC;
XX
XX WPI; 2004-099381/10.
XX
XX P-PSDB; ADJ93358.
XX
XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,
XX useful for preventing, treating or ameliorating a medical condition, e.g.
XX aberrant cellular proliferation, reproductive disorders or testicular
XX disorders.
XX
XX Claim 1; SEQ ID NO 1; 343pp; English.
XX
XX This invention relates to a novel testis-specific tubulin tyrosine-ligase
XX -like polypeptide, designated the BGS-42 polypeptide. The invention may
XX be useful for the development of compounds with a cytostatic, respiratory
XX -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,

CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a cDNA
CC which encodes the human BGS-42 protein of the invention.

XX SQ Sequence 1838 BP; 381 A; 576 C; 583 G; 298 T; 0 U; 0 Other;

Query Match 12.2%; Score 272.4; DB 12; Length 1838;
Best Local Similarity 99.6%; Pred. No. 2.7e-49;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCATCCGCTATGAGGCGGGAAGTGGGCGGTGGGGAGCC 1965
Db 1 TGGGAGCAGGCTGGGCCCCATCCGCTATGAGGCGGGAAGTGGGCGGTGGGGAGCC 60

Qy 1966 TCCGTGGCCCTGGGCCCCATCCGCTATGAGGCGGGAAGTGGGCGGTGGGGAGCC 2025
Db 61 TCCGTGGCCCTGGGCCCCATCCGCTATGAGGCGGGAAGTGGGCGGTGGGGAGCC 120

Qy 2026 GTGCTGTCTTTTTCAGAGACTTCGCGGCGCACCATGGCTCCGACATCCTCAAGTGGGTGG 2085
Db 121 GTGCTGTCTTTTTCAGAGACTTCGCGGCGCACCATGGCTCCGACATCCTCAAGTGGGTGG 180

Qy 2086 TCAGCCACACAGAGTGCAGCAGGAGCAGCAGAGCAAGCCAGGACAGAGGAGGAGG 2145
Db 181 TCAGCCACACAGAGTGCAGCAGGAGCAGCAGAGCAAGCCAGGACAGAGGAGGAGG 240

Qy 2146 CCGGGAGCAGGACCTGAGCAGGAGGCAAGTGC 2179
Db 241 CCGGGAGCAGGACCTGAGCAGGAGGCAAGTGC 274

RESULT 6

ADJ93361

ID ADJ93361 standard; DNA; 1939 BP.

XX AC ADJ93361;

XX AC ADJ93361;

DT 06-MAY-2004 (first entry)

XX Human BGS-42 protein-related DNA clone A SeqID9.

DE testis-specific tubulin tyrosine-ligase-like polypeptide;
XX BGS-42 polypeptide; cytosolic; respiratory-Gen; gastrointestinal-Gen;
XX neuroprotective; endocrine-Gen; antiinflammatory; anabolic; hypertensive;
XX osteopathic; nootropic; antiparkinsonian; antiarthritic; antiasthmatic;
XX anti-HIV; antibacterial; immunosuppressive; antiseborrheic;
XX dermatological; tyrosine ligase modulator; gene therapy; tubulin ligase;
XX tubulin-carboxypeptidase; cellular proliferation; reproductive disorder;
XX testicular disorder; testicular cancer; pulmonary disorder; lung cancer;
XX gastrointestinal disorder; colon cancer; stomach cancer; neural disorder;
XX brain cancer; liver cancer; proliferative condition; testis; lung;
XX small intestine; brain; lymph tissue; infertility; Cushing's syndrome;
XX emphysema; pneumonia; Addison's disease; acromegaly; Alzheimer's disease;
XX Parkinson's disease; immunological disorder; arthritis; asthma; AIDS;

KW sepsis; acne; Sjogren's disease; scleroderma; human; ds.

XX Homo sapiens.

OS Homo sapiens.

PN W02004005487-A2.

XX 15-JAN-2004.

XX 09-JUL-2003; 2003WO-US021605.

XX 09-JUL-2002; 2002US-0394725P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Feder JN, Wu S, Nelson TC;

XX WPI; 2004-099381/10.

XX New testis-specific tubulin tyrosine-ligase-like BGS-42 polypeptide,

PT useful for preventing, treating or ameliorating a medical condition, e.g.

PT aberrant cellular proliferation, reproductive disorders or testicular

PT disorders.

XX Example 4; SEQ ID NO 9; 343pp; English.

XX

CC This invention relates to a novel testis-specific tubulin tyrosine-ligase
CC -like polypeptide, designated the BGS-42 polypeptide. The invention may
CC be useful for the development of compounds with a cytostatic, respiratory
CC -Gen, gastrointestinal-Gen, neuroprotective, endocrine-Gen,
CC antiinflammatory, anabolic, hypertensive, osteopathic, nootropic,
CC antiparkinsonian, antiarthritic, antiasthmatic, anti-HIV, antibacterial,
CC immunosuppressive, antiseborrheic or dermatological activity acting as
CC tyrosine ligase modulators. In addition, the disclosed sequences may be
CC useful for gene therapy. The BGS-42 polypeptide or polynucleotide can be
CC used for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, and for preventing, treating or
CC ameliorating a medical condition, such as a disorder related to aberrant
CC tubulin ligase activity, a disorder related to aberrant tubulin-
CC carboxypeptidase activity, aberrant cellular proliferation, reproductive
CC disorders, testicular disorders, testicular cancer, pulmonary disorders,
CC lung cancer, gastrointestinal disorders, colon cancer, stomach cancer,
CC neural disorders, brain cancer, liver cancer, or proliferative condition
CC of the testis, lung, small intestine, brain or lymph tissue. The BGS-42
CC polypeptide, polynucleotide, or their modulators are also useful for
CC treating infertility, Cushing's syndrome, emphysema, pneumonia, Addison's
CC disease, acromegaly, Alzheimer's disease, or Parkinson's disease. The BGS
CC -42 polypeptide can be used as a preventive agent for immunological
CC disorders including arthritis, asthma, AIDS, sepsis, acne, Sjogren's
CC disease or scleroderma. The antibodies may be used to purify, detect and
CC target the BGS-42 polypeptides. The present sequence is that of a DNA
CC clone sequence which is related to the invention.

XX SQ Sequence 1939 BP; 421 A; 592 C; 589 G; 337 T; 0 U; 0 Other;

Query Match 12.2%; Score 272.4; DB 12; Length 1939;
Best Local Similarity 99.6%; Pred. No. 2.7e-49;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1906 TGGGAGCAGGCTGGGCCCCATCCGCTATGAGGCGGGAAGTGGGCGGTGGGGAGCC 1965
Db 1 TGGGAGCAGGCTGGGCCCCATCCGCTATGAGGCGGGAAGTGGGCGGTGGGGAGCC 60

Qy 1966 TCCGTGGCCCTGGGCCCCATCCGCTATGAGGCGGGAAGTGGGCGGTGGGGAGCC 2025
Db 61 TCCGTGGCCCTGGGCCCCATCCGCTATGAGGCGGGAAGTGGGCGGTGGGGAGCC 120

Qy 2026 GTGCTGTCTTTTTCAGAGACTTCGCGGCGCACCATGGCTCCGACATCCTCAAGTGGGTGG 2085
Db 121 GTGCTGTCTTTTTCAGAGACTTCGCGGCGCACCATGGCTCCGACATCCTCAAGTGGGTGG 180

Qy 2086 TCAGCCACACAGAGTGCAGCAGGAGCAGCAGAGCAAGCCAGGACAGAGGAGGAGG 2145
Db 181 TCAGCCACACAGAGTGCAGCAGGAGCAGCAGAGCAAGCCAGGACAGAGGAGGAGG 240

CC dementia and various cancers. This sequence corresponds to a nucleotide
CC sequence of the invention.
SQ Sequence 5282 BP; 1209 A; 1482 C; 1362 G; 1229 T; 0 U; 0 Other;
Query Match 3.2%; Score 71.6; DB 12; Length 5282;
Best Local Similarity 63.2%; Pred. No. 1.6e-05;
Matches 110; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 96 TGAATGCTGGGCATCAATAAAGACTTTTCTTCCAGTCTAGGTTGGTAAATAATGAGA 155
Db 6 TGAATTTGATGACTTAGTGAACAACATGCTCTGATGTCGCGATGTCAGATGAGA 65
Qy 156 TGCCTGACCTCTCTGGACCATCAAGAGGACATCATTTGACTATFCACAGCTGACCTACG 215
Db 66 TCCCTACTTCTATCTGGACCACTCGCGGGATGTCGAGTGTCTCTCTCCAAGG 125
Qy 216 ACAGATGCTGAACCACTAGCAAGAAGACAGCTCTCTTACCACCAAGGTAGGCC 269
Db 126 ATCAGATGATAAACCACTAGCCCGGGCTGGCTCTTTTACCACAAAGGTGGGTC 179
RESULT 9
ABQ97805
ID ABQ97805 standard; cDNA; 452 BP.
AC ABQ97805;
XX
XX 30-OCT-2002 (first entry)
DE Mouse ES cell related cDNA SEQ ID NO 1073.
KW Mouse; ES cell; gene trapped sequence; GTS; gene expression;
KW development disorder; cell differentiation disorder; gene; ss.
XX
OS Mus sp.
XX
XX US2002081668-A1.
XX
XX 27-JUN-2002.
XX
XX 30-NOV-2000; 2000US-00728446.
XX
XX 20-NOV-1998; 98US-0109302P.
XX 01-DEC-1999; 99US-0168270P.
XX
XX (FRIE/) FRIEDRICH G.
XX (ZAMB/) ZAMBROWICZ B.
XX (SAND/) SANDS A T.
XX
XX Friedrich G, Zambrowicz B, Sands AT;
PI
XX
XX WPI; 2002-626541/67.
XX
XX Novel murine polynucleotides that individually identify novel genes into
PT which a retroviral gene trap vector has been integrated, useful in
PT genomic analysis and in discovery, development of therapeutic and
PT diagnostic agents.
XX
XX Claim 2; SEQ ID NO 1073; 29pp + Sequence Listing; English.
XX
XX The invention relates to isolated murine polynucleotides (I) comprising a
CC contiguous stretch of at least about 60 nucleotides of a sequence
CC (ABQ97805-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences
CC (GTSs). The novel genes can be used in a process to identify novel
CC polynucleotide sequences by comparing them to the novel gene sequences.
CC The novel genes and cells are useful in functional genomic analysis and
CC in the discovery and development of new therapeutic and diagnostic agents
CC and methods. (I) is useful for identifying the coding regions of the
CC murine genome, to isolate cDNAs, genomic clones or full-length
CC genes/polynucleotides or homologues, heterologues, paralogues or
CC orthologues that are capable of hybridising to one or more of the GTSs
CC under stringent conditions. (I) can be incorporated into a phage display

CC system that can be used to screen for proteins or other ligands, that are
CC capable of binding an amino acid sequence encoded by an oligonucleotide
CC or polynucleotide sequence in at least one of the GTS sequences. (I) is
CC useful in arrays, such as gene chips, to identify and characterise
CC temporal and tissue specific gene expression, to identify the gene of
CC interest from many sources and for genetic manipulations such as
CC antisense inhibition and gene targeting. Decreasing the level of
CC expression of (I) and/or down regulating the activity of peptides or
CC proteins encoded by (I) is useful for treating development and cell
CC differentiation disorders. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20020081668
XX
SQ Sequence 452 BP; 94 A; 117 C; 115 G; 98 T; 0 U; 28 Other;
Query Match 3.1%; Score 69.2; DB 6; Length 452;
Best Local Similarity 68.8%; Pred. No. 3.1e-05;
Matches 95; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
Qy 132 AGTCTAGGTTGGTAAAAATGAGATCCGTACCTCTCTCGACCATCAAGAGGGACATCA 191
Db 6 AGTCCCGCATGGTTCGAATGAGACCCCTACTCTCATCTGGACCACTCGCGGGATGTC 65
Qy 192 TTGACTATACAGCTGACCTACGACAGATGCTGAACCACTACGGAAGACAGCTCTCT 251
Db 66 TGGATTGTGCTTCTTTCCAGGATCAGATGATAAACCATTTATGCCCGTCGAGGCTCT 125
Qy 252 TCACCAACAGGTGAGCC 269
Db 126 TCACCAACAGGTGAGCC 143
RESULT 10
AAS72428
ID AAS72428 standard; cDNA; 1042 BP.
XX
XX AAS72428;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #8232.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX P-PSDB; ABG08241.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 8232; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1042 BP; 227 A; 332 C; 288 G; 195 T; 0 U; 0 Other;

Query Match 3.0%; Score 67.8; DB 5; Length 1042;
Best Local Similarity 62.9%; Pred. No. 7.5e-05;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 96 TGAATGCTGGGCATCAATAAAGACTTTTCTTCCAGCTCTAGTTGGTAAATAATGAGA 155
DB 707 TGAATTTTGATGACCTAGATGGAACACATGCTCTGATGTCCGCGATGGTCCAGAAATGAGA 766
QY 156 TGCCGTACTCTCTGGACCATCAAGAGGACATCATTTGACCTATCAGCTGATGCTCTCTCTCAAGG 826
DB 767 TCCCTACTTCACTTGGACCACTCGCGGGATGCTCGACTGCTCTCTCTCAAGG 826
QY 216 ACCAGATGCTGAACCACTACGCAAGACAGCGCTCTCTCAACCAAG 262
DB 827 ATCAGATGATAAACCACTACGCGGGTGGCTCTCTTACCAACAAG 873

RESULT 11
AAS69383
ID AAS69383 standard; cDNA; 2250 BP.
AC AAS69383;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #5187.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI
DR WPI; 2001-639362/73.
DR P-PSDB; ABG05196.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

PT biodiversity.
XX
PS Claim 1; SEQ ID NO 5187; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 2250 BP; 699 A; 558 C; 620 G; 373 T; 0 U; 0 Other;

Query Match 3.0%; Score 67.8; DB 5; Length 2250;
Best Local Similarity 62.9%; Pred. No. 8.9e-05;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 96 TGAATGCTGGGCATCAATAAAGACTTTTCTTCCAGCTCTAGTTGGTAAATAATGAGA 155
DB 1427 TGAATTTTGATGACCTAGATGGAACACATGCTCTGATGTCCGCGATGGTCCAGAAATGAGA 1486
QY 156 TGCCGTACTCTCTGGACCATCAAGAGGACATCATTTGACCTATCAGCTGATGCTCTCTCTCAAGG 215
DB 1487 TCCCTACTTCACTTGGACCACTCGCGGGATGCTCGACTGCTCTCTCTCTCAAGG 1546

RESULT 12
ADC86736
ID ADC86736 standard; DNA; 5452 BP.
XX
AC ADC86736;
XX
DT 01-JAN-2004 (first entry)
DE Human GPCR gene SEQ ID NO:1189.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX Homo sapiens.
XX
PN EP1270724-A2.
XX
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX

QY 1899 GGGGGTGTGGGAGCAGGCTGGCCCATCGGCTATGAGGGCGGAGTGGGGGGTTG 1958
|||||
Db 260 GGGGGGGGAGGG 319
|||||

QY 1959 GGG 1961
|||

Db 320 GGG 322
|||

RESULT 14

AAZ17263/c

ID AAZ17263 standard; cDNA; 1337 BP.

XX AC AAZ17263;

XX DT 12-OCT-1999 (first entry)

XX DE Human gene expression product cDNA sequence SEQ ID NO:4735.

XX KW Human; gene; gene expression product; diagnosis; therapy; probe;
XX KW detection; mapping; tissue typing; profiling; forensic; cancer;
XX KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX OS Homo sapiens.

XX PN WO9938972-A2.

XX PD 05-AUG-1999.

XX PF 28-JAN-1999; 99WO-US001619.

XX PR 28-JAN-1998; 98US-0072910P.

XX PR 24-FEB-1998; 98US-0075954P.

XX PR 31-MAR-1998; 98US-0080114P.

XX PR 03-APR-1998; 98US-0080515P.

XX PR 03-APR-1998; 98US-0080666P.

XX PR 21-OCT-1998; 98US-0105234P.

XX PR 28-OCT-1998; 98US-0105877P.

XX PA (CHIR) CHIRON CORP.

XX PA (HYSE-) HYSEQ INC.

XX PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

XX PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

XX PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

XX PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;

XX DR WPI; 1999-494092/41.

XX PT Novel human genes and their expression products which are differentially

XX PT expressed in different cell types.

XX PS Claim 1; Page 2250-2251; 2479pp; English.

XX CC The present invention describes a library of human polynucleotides

XX CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is

XX CC a method of detecting differentially expressed genes correlated with the

XX CC cancerous state of a mammalian cell, comprising detecting at least one

XX CC differentially expressed gene product in a test sample from a cell

XX CC suspected of being cancerous, where the gene product is encoded by one of

XX CC the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The

XX CC polynucleotides can be used as a source of primers and probes, which can

XX CC be used for a variety of purposes, e.g. detection of expression levels,

XX CC mapping, tissue typing or profiling, forensics, genetic analysis and

CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists

XX SQ Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 0 U; 669 Other;

Query Match 2.6%; Score 58; DB 2; Length 1337;

Best Local Similarity 26.7%; Pred. No. 0.011;

Matches 188; Conservative 0; Mismatches 512; Indels 3; Gaps 1;

QY 1273 GCTGAGGGGGTGTCTCTGGGGCCCTGGGGAGCCGTATATCTGGGAGGGGAGGATCG 1332

Db 1183 GNG 1124

QY 1333 CCTAGGGGTGTGTGTGTGTCCTCCCATAGCCAGGCAAGTCTGAGGTGAGCGGCCCTCCT 1392

Db 1123 NGNNNGGGGNG 1064

QY 1393 CGCTTTGTGAGCGAGGAGAGCTGAGCTGAGCTGCCCTCTCTGTAAACACCCACTGTC 1452

Db 1063 NNNNNNG 1004

QY 1453 TCTGAGATCGGGCTGTGCGTGAACATCGGAGCCTGCCCTCGTACGTCGCCCAACCC 1512

Db 1003 GNNNG 944

QY 1513 CGACTCTCTTCCACGCTGTACAGCTCTGACCGAGAGTGAGCAGAGAGTTCCT 1572

Db 943 NGNNNG 884

QY 1573 GGTTAAGTGAAGAGCGGAGAGGGCCCCCAGTGTCTGACGAGAGAGGCTTTAGAAAGA 1632

Db 883 GNNNG 824

QY 1633 TCCCCCTGTCTGAGACAGACTGATGGCGAGCGTCTGAGGATAGAGACCGGGGAGAG 1692

Db 823 GNGNGNGGGGNGNGNGNGNGNGNGNGNGNGNGNGNGNGNGNGNGNG 764

QY 1693 GCCTCCCATGTGTCATGGTCAATGGCA---GTACAGAGGCCAGGGCCCCCGGAGGGAGGGC 1749

Db 763 GNNNGNGGGGNGGGGNGGGGNGGGGNGGGGNGGGGNGGGGNGGGGNGGGG 704

QY 1750 GGGCAGTCAAGAGTGTGTGTCTGTCTAGGCTTGTGGACAGCGCGGAGCAGCTGG 1809

Db 703 NGGGGGGGGNGGNGNGNGNGNGGGGGGNGGNGGGGGGGGGGGGGGGGGG 644

QY 1810 GGCAGAGCGCGCTCCGTCGTGCGAGCGCCAGGGGTGCAGTGGACGCGCGCAGTCACACA 1869

Db 643 NGGGGGGGGGGNGGNGGNGGNGGGGGGNGGGGGGNGGGGNGGGGNGGGG 584

QY 1870 GACACTGAGGGAGAGAGCGCAGGACAGACTGGGGGTGTGGGAGCAGGCTTGGCCCCATCG 1929

Db 583 GNGGNGGGGGGNGGNGGNGGNGGGGNGGNGGNGGNGGNGGNGGNGGNGG 524

QY 1930 GCTATGAGGGCGGAGTGGGGCGGGTGGGAGCCTCGGTGG 1972

Db 523 GNGGGGGGGGGGGGNGGGGNGGGGNGGGGNGGGGNGGGGNGGGGNGGGG 481

RESULT 15

ACN54596

ID ACN54596 standard; cDNA; 588 BP.

XX AC ACN54596;

XX DT 02-DEC-2004 (first entry)

XX DE Cotton androecium tissue EST Clone ID: LIB3828-025-Q6-K6-D5, SEQ:9377.

XX KW Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;

XX KW variety Nucotton33B; library LIB3828; molecular tag; molecular marker;

XX KW genetic mapping; molecular mapping; seed germination; plant growth;

XX KW plant quality; plant yield; plant breeding; tissue printing; ss.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2005, 11:08:28 ; Search time 8612.22 Seconds
(without alignment)
9904.766 Million cell updates/sec

Title: US-10-615-659-27

Perfect score: 2241

Sequence: 1 tgcctgcagcagcccgccg.....gcctccacagggccctccc 2241

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gse1:*

9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	137.4	6.1	1259	9	AY415398	Homo sapi
2	135.8	6.1	1257	9	AY415399	Pan trogl
3	129	5.8	129	7	H55115	CHR220054 C
4	111.8	5.0	213	7	H55109	CHR220048 C
5	97	4.3	2969	3	AK030151	Mus muscu
6	93.8	4.2	487	8	AZ248094	RPT-23-4
7	93.8	4.2	2720	3	AK029745	Mus muscu
8	83.2	3.7	1218	9	CU081011	CH216-160
9	76.4	3.4	467	2	BF903338	PMI-MT019
10	74	3.3	623	6	BY745669	BY745669
11	74	3.3	2520	3	AK080321	Mus muscu
12	73.4	3.3	1174	6	CU504184	CDR66-E10
13	73.2	3.3	1585	8	CC294035	CH261-82A
14	71.6	3.2	3328	3	BC028169	Homo sapi
15	71.4	3.2	1040	9	CNS030AX	AL253122 Tetraodon
16	70.8	3.2	560	9	CG564652	OST192793
17	70.8	3.2	741	7	CNS26715	UI-M-H80-
18	70.8	3.2	1451	3	AK029462	Mus muscu
19	70.6	3.2	582	5	BP292995	BP292995
20	70.6	3.2	1315	9	AG279688	Mus muscu
21	70.6	3.2	1909	9	CU078604	CH216-151
22	70	3.1	252	9	CG546562	OST146147
23	70	3.1	925	9	CNS0091P	Drosophila
24	69.8	3.1	581	5	BP357921	BP357921

25	69.6	3.1	1069	9	CU118121	ISB1-70B1
26	69.4	3.1	1137	9	AG078502	Pan trogl
27	69.2	3.1	251	9	CG592880	OST49456
28	69.2	3.1	251	9	CG599359	OST266453
29	69.2	3.1	378	9	CG624680	OST329018
30	69.2	3.1	452	9	CG532512	OST116474
31	69	3.1	910	8	BH162948	ENTR49TR
32	68.8	3.1	525	7	CF916454	B0995D07-
33	68.6	3.1	519	2	AW656035	108059 MA
34	68.6	3.1	573	2	AW656034	108058 MA
35	68.6	3.1	825	7	CK770333	958489 MA
36	68.4	3.1	821	9	CNS049PN	Tetraodon
37	68.4	3.1	891	9	CU477366	SAIL_271
38	68.4	3.1	1038	9	CU478560	SAIL_294-
39	68.2	3.0	1340	9	AG430438	Mus muscu
40	68.2	3.0	1405	9	CU509298	SAIL_811
41	68	3.0	1690	9	AG435238	Mus muscu
42	68	3.0	1767	8	CC294031	CH261-82A
43	67.8	3.0	1153	8	CC252982	CH261-187
44	67.8	3.0	1269	5	BN914457	AGENCOURT
45	67.8	3.0	1523	9	AG448267	Mus muscu

ALIGNMENTS

RESULT 1
AY415398
LOCUS
DEFINITION Homo sapiens HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY415398
VERSION AY415398.1 GI:39771357
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1259)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1259)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source 1..1259
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1259
/locus_tag="HCM5544"

Query Match 6.1%; Score 137.4; DB 9; Length 1259;
Best Local Similarity 99.3%; Pred. No. 2.6e-19;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2041 AAGACTTCCGCGCGCAGCATCCAGCATCCCTCAAGTGGTGGTTCAGCCAGCAGCT 2100
Db 1 AAGACTTCCGCGCGCAGCATCCAGCATCCCTCAAGTGGTGGTTCAGCCAGCAGCT 60

Qy	2101	GCAGCAGGACGACGAGAAGCAAGCCCAAGGACCGAGGGAGGAGGCCGGGAGCGACGACC	2160
Ds	61	GCAGCAGGACGACGAGAAGCAAGCCCAAGGACCGAGGGAGGAGGCCGGGAGCGACGACC	120
Qy	2161	TGAGCAGCAGCGAAGGTGC	2179
Ds	121	TGAGCAGCAGCGAAGGTGC	139
RESULT 2			
AY415399			
LOCUS		1257 bp DNA linear GSS 12-DEC-2003	
DEFINITION		Pan troglodytes HCM5544 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.	
ACCESSION		AY415399	
VERSION		AY415399.1 GI:39771358	
KEYWORDS		GSS.	
SOURCE		Pan troglodytes (chimpanzee)	
ORGANISM		Pan troglodytes	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan. 1 (bases 1 to 1257) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios Science 302 (5652), 1960-1963 (2003) 14671302	
REFERENCE		2 (bases 1 to 1257) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M. Direct Submission Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA This sequence was made by sequencing genomic exons and ordering them based on alignment. Location/Qualifiers 1..1257 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" <1..->1257 /locus_tag="HCM5544"	
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
gene			
ORIGIN			
Query Match		6.1%; Score 135.8; DB 9; Length 1257;	
Best Local Similarity		98.6%; Pred. No. 5.9e-19;	
Matches 137; Conservative		0; Mismatches 2; Indels 0; Gaps 0;	
Qy	2041	AAGACTTCGGCGCACCATGGATCCAGCATCTCTCAAGTGGTGTCAGCCACAGAGCT	2100
Ds	1	AAGACTTCGGCGCACCATGGATCCAGCATCTCTCAAGTGGTGTCAGCCACAGAGCT	60
Qy	2101	GCAGCAGGACGACGAGAAGCAAGCCCAAGGACCGAGGGAGGAGGCCGGGAGCGACGACC	2160
Ds	61	GCAGCAGGACGACGAGAAGCAAGCCCAAGGACCGAGGGAGGAGGCCGGGAGCGACGACC	120
Qy	2161	TGAGCAGCAGCGAAGGTGC	2179
Ds	121	TGAGCAGCAGCGAAGGTGC	139
RESULT 3			
H55115			
LOCUS		129 bp mRNA linear EST 07-DEC-1995	
DEFINITION		CHR220054 Chromosome 22 exon Homo sapiens cDNA clone C22_74 5', mRNA sequence.	
ACCESSION		H55115	

8593609
 COMMENT Contact: Buckler AJ
 Molecular Neurogenetics Unit
 Massachusetts General Hospital
 Building 149, 13th St., Charlestown MA 02129
 Tel: 6177249616
 Fax: 6177265736
 Email: buckler@helix.mgh.harvard.edu
 Seq primer: 73.

FEATURES
 source Location/Qualifiers
 1..213
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="C22-67"
 /lab_host="E. coli DH5a"
 /clone_lib="Chromosome 22 exon"
 /note="Vector: pBluescriptIIKS+; Site 1: Sal I; Site 2: Bam HI (destroyed); Exons were isolated from human chromosome 22 specific cosmids using a modification of the method of exon amplification (Proc. Natl. Acad. Sci. USA 88:4005-4009, 1991). Amplified exons were digested with Sal I and Bgl II and subsequently cloned into pBluescriptIIKS+ at the Sal I and Bam HI sites."

ORIGIN
 Query Match 5.0%; Score 111.8; DB 7; Length 213;
 Best Local Similarity 98.3%; Pred. No. 8.9e-14;
 Matches 113; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1460 ATCGGGCTGTGCTGAACATCGGAGCTGCGCTACGTCCCGCCACCCGACTCC 1519
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ATCGGGCTGTGCTGAACATCGGAGCTGCGCTACGTCCCGCCACCCGACTCC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 1520 TTCTTCCACGCTGTACAGCTCTGACCGAGAGTGAGCAGAGTCTCTGG 1574
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 TTCTTCCACGCTGTACAGCTCTGACCGAGAGTGAGCAGAGTCTCTGG 115
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 5
 AK030151
 LOCUS AK030151 2969 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933401B17 product:hypothetical tubulin-tyrosine ligase containing protein, full insert sequence.

ACCESSION AK030151.1 GI:26326134
 VERSION AK030151
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 Carninci, P., and Hayashizaki, Y.
 2 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

REFERENCE 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

REFERENCE 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

6 (bases 1 to 2969)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tanagi, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedic Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 Location/Qualifiers
 1..2969
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:4933401B17"
 /db_xref="taxon:10090"
 /clone="4933401B17"
 /sex="male"
 /tissue_type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 490..2835
 /note="unnamed protein product; hypothetical tubulin-tyrosine ligase containing protein (InterPro|IPR004344, evidence: InterPro) putative"
 /codon_start=1
 /protein_id="BAC26811.1"
 /db_xref="GI:26326135"
 /translation="MENGKKSLSDGDHKEENKLGKIPDLSSPKLDYKIA RQUTEAKERKFIYGHYPVIRATLRKGNVKKFNFPKALQNLGSEKSAETKE KIGLALRFDIDHVMRLVNEIPYLLWTIKRDVYHSLTCDMLNHYGKTASFTT KIGLNLRSPLPWYQANPNTFFRCYGLCTSEKQFLDFFRTVAASILKVVVLHQ DSVSKVKGSKSEAKNSDPSPKDPENPDLPKLSLQVVDFTVCQVAYLGOLEHE DNVSEASTALSEEENDLTQOYLLVHGNASITDSKVPACQALLSKISSVNPOT EIDGIRNIWIKPAKSRGRDVCMDRVENILSLVAADSTQTKDNKVVQYK1ETPML

FEATURES
 source
 1..2969
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:4933401B17"
 /db_xref="taxon:10090"
 /clone="4933401B17"
 /sex="male"
 /tissue_type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 490..2835
 /note="unnamed protein product; hypothetical tubulin-tyrosine ligase containing protein (InterPro|IPR004344, evidence: InterPro) putative"
 /codon_start=1
 /protein_id="BAC26811.1"
 /db_xref="GI:26326135"
 /translation="MENGKKSLSDGDHKEENKLGKIPDLSSPKLDYKIA RQUTEAKERKFIYGHYPVIRATLRKGNVKKFNFPKALQNLGSEKSAETKE KIGLALRFDIDHVMRLVNEIPYLLWTIKRDVYHSLTCDMLNHYGKTASFTT KIGLNLRSPLPWYQANPNTFFRCYGLCTSEKQFLDFFRTVAASILKVVVLHQ DSVSKVKGSKSEAKNSDPSPKDPENPDLPKLSLQVVDFTVCQVAYLGOLEHE DNVSEASTALSEEENDLTQOYLLVHGNASITDSKVPACQALLSKISSVNPOT EIDGIRNIWIKPAKSRGRDVCMDRVENILSLVAADSTQTKDNKVVQYK1ETPML

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsf.riken.jp, URL: <http://genome.gsf.riken.jp/>
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site (<http://genome.gsf.riken.go.jp>) for
further details.

FEATURES

Location/Qualifiers
1..623
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630053H17"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 3 days neonate
thymus"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTCTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5' GAGAGAGAGATCTCGAGTAAATTAATTAATTCCTCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

ORIGIN

Query Match 3.3%; Score 74; DB 6; Length 623;
Best Local Similarity 69.2%; Pred. No. 2e-05;
Matches 101; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
124 TTCTTGGCAGTCTAGGTGGTAAATATGAGATCCGCTACCTCTCTGGACCATCAAGAG 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 TTCTTCCAGTCCCGCATGGTTCCGAATGAGACCCCTACCTCTCTCTGGACCATCTCGCG 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 184 GGACATCATTTGACTATCACAGCCTGACCTAGCAGATGCTGAACCACTACGCAAGAC 243
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 GGATGCTGATGTGCTCTTCCTTTTCCAGGATCAGATGATAAACCATTATGCCGTGC 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 244 AGCTCTCTTCCACCAACCAAGGTGAGCC 269
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 AGGTCTCTTCCACCAACCAAGGTGAGCC 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AK080321
LOCUS
DEFINITION

Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
library, clone:A630053H17 product:HOTT1 PROTEIN homolog [Homo
sapiens], full insert sequence.
AK080321
AK080321.1 GI:26348476
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
1

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
REFERENCE
2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
REFERENCE
3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
REFERENCE
4

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
REFERENCE
5

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
REFERENCE
6

(bases 1 to 2520)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, I., Kasukawa, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Konno, H., Kouda, M.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuura, T., Miyazaki, R., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

FEATURES
source
1. .2520
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630053H17"
/db_xref="taxon:10090"
/clone="A630053H17"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
84. .2198
/note="unnamed protein product; HOTTL PROTEIN homolog [Homo sapiens] [SPTR|Q5UI99, evidence: PASTY, 85.8%ID, 93.4%length, match=821]
putative"
/codon_start=1
/protein_id="BAC37878.1"
/db_xref="GI:26348477"
/translation="MDTSSPFYAAGGCEASPPSLPLVSISSPSLQSRWVRNETPVLWT
TRDVLDCRFLSKQDMLNHYVARAGSFYTKVGLCLNLRNLPWFDEADSFPRCYRLG
AEDKXAFIDFWLTAARNVLKLVLEKESQSIQAREEAPEDTPQPKBKLLVT
VSDVFQALSAQQLHNSVLPQIDMEGRINIVKFGAKSRGRGIMCMRLDMLKLV
YLSVQVQRCDEIIQHLNVVPLDIFGTGKDFLRQFLVFNWPLTVFYRDSYTRFTQ
DCNPLMKGKWTQVIERPLIFGTGKDFLRQFLVFNWPLTVFYRDSYTRFTQ
PFSLKNLDNSVHLNNSIORLEASCHRHPLPDDNWSORFOAHLQEVDAFKWSS
VIVPGKAAVIHALQTSQDNVQCRKASFLYGDVFGEDFQPLWLEINASTPMAST
AVTARLCAGVQADTLRVVDRDLRSCDTGAFELITYKQPAVEVFQIVGIRLLVEGSTI
KKPVPVGHRTGTVRSLPHLLTQGGESKSGSPHRSASRKNRAESLHETKEPEP
AAVASGKGKAPFPFSLHSAWLPSPRVHRPQGRVLRQLQDLVGSKALSTGKA
LMTLPTAKVLMSFPFPHDLKLPMSLKPQGVPELCCTTWRVLSGGIGEGHRQAA
PRSSAPGKGLSSTPCKSTET"

CDS
[Homo sapiens] [SPTR|Q5UI99, evidence: PASTY, 85.8%ID, 93.4%length, match=821]
putative"

ORIGIN
Query Match 3.3%; Score 74; DB 3; Length 2520;
Best Local Similarity 69.2%; Pred. No. 2.4e-05;
Matches 101; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 124 TTCTTCCAGTCTAGGTTGGTAAATAATGAGATCCGTACTCTCTGGACCATCAAG 183
DB 164 TTCTCTCCAGTCCGCGATGGTTGAAATGAGACCCCTTACTCTATCTGGACCAT 223
QY 184 GGACATCATTTGACTATCACACGCTGACCTACGACCATGCTGAACCACTACGCAAGAC 243
DB 224 GGATGCTGATTTGCTTCTCTTTCACAGGATCAGATGATAACCATTTATGCCGTGC 283
QY 244 AGGCTCTCTTCCACCAAGGTGAGCC 269
DB 284 AGGCTCTTCCACCAAGGTGAGCC 309

RESULT 12
CD504184

LOCUS CD504184 1174 bp mRNA linear EST 12-JUN-2003
DEFINITION CDA66-E10.X1d-t SRGC-CDA Gasterosteus aculeatus cDNA clone
CDA66-E10 5', mRNA sequence.
ACCESSION CD504184
VERSION CD504184.1 GI:31434752
KEYWORDS EST.
SOURCE Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1174)
Kingley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingseley@cmgm.stanford.edu
Plate: 66
High quality sequence stop: 289.

FEATURES
source
1. .1174
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA66-E10"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

ORIGIN

Query Match 3.3%; Score 73.4; DB 6; Length 1174;
Best Local Similarity 28.7%; Pred. No. 3e-05;
Matches 206; Conservative 0; Mismatches 513; Indels 0; Gaps 0;
QY 1483 GAGCCCTGCCCTGGTACGTCCCGGCCAACCCGACTCTCTCCACGCTGTACAGCCT 1542
DB 204 GAGCCAGTACCGACATCGCCGACTCACCGGCGACTCGGGCATCAGGTCTGCCGGTGTG 263
QY 1543 CTGCACCGAGGTGAGCAGCAGGAGTTCCTGGGTAAGTGAGGAGACGCGAGAGGGCCCCA 1602
DB 264 GGTGGAGGAGGAGGAGGANNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 323
QY 1603 GTCTGTGACGAGAGGCTTCTAGAAAGATCCCCCTGGTGTGACGACACTGATGGGG 1662
DB 324 NNN 383
QY 1663 CAGGCTCAGGATAGAGGACCGGGAGAGCCCTCCCATGCTCATGTCATGCAGTACA 1722
DB 384 NNNNGGGNNNNNNNNNNNNNGNNGNNGNNGNNGNNGNNGNNGNNGGNGG 443

[illegible]

Query Match 3.3%; Score 73.2; DB 8; Length 1585;
Best Local Similarity 41.8%; Pred. No. 3.4e-05;
Matches 396; Conservative 0; Mismatches 551; Indels 0; Gaps 0;

QY 1276 GAGGGGTTCTGGGCCCTTGGGGACCGATATATCTGGGGAGGGGAGGATCCGCC 1335
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1317 GGAAGGGGNGGG 1258
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1336 TAGGGGTGCTGTCGTGTGCCCATAGCAGGCAAGTCTGAGGTGGACGGCCGTCCTCGC 1395
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1257 TTGGGNGGG 1198
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1396 CTTGTGAGCGAGGCAGAGCTGAGCTGAGCTGCCCCCTCTGCTTAACACGCCACTGTCTCT 1455
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1197 GGSGGCGGG 1138
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1456 GGAATCGGGCTGTGCTGTAACAATGGAGCCTGCCCTGTGTAGTCTCCGGCCAACCCGA 1515
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1137 GGG 1078
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1516 CTCCTCTTCCCACGCTGTACAGCCTCTGCACCGAGAGTGACAGCAGAGTTCTCTGGG 1575
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1077 GGG 1018
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1576 TAAGTGAGGAGACGGCAGAGGGCCCCAGTCTGTACAGAGAGAGCTTCTAGAAAGATCC 1635
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1017 GGTGG 958
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1636 CCTGTGTGTGACACAGACTGATGGGCGAGGTCTTAGAGATAGAGACCAGGAGAGGCC 1695
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 957 CGGGGAGGGGGGGAGCGCGTGTGGGACCGGGAGGTGGGGGGGGGGGGGGGGGGGG 898
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1696 TCCCATGTCTATGTCATGGCAGTACAGAGGCCAGGGCCCCGGGAGGAGGGCGGCAG 1755
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 897 GGGCAGGGGGGGGGGGTGGGGGGGGGGTGGGGGGGGGGTGGGAGGGGTGGGGGGGGCGA 838
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1756 TCAAGAGGTGTGTGTCTGTGTAGGCTGTGGAAGCGCGCGAGCACTCGGGGCAAG 1815
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 837 GGGGTGG 778
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1816 GCCGGTCCGTGTCCGAGGCCAGGGGTGCAGCTGGAAGGCCGCACTCACAGACACT 1875
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 777 GGG 718
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1876 GCAGGAGAAGGCACGACAGACTGGGGGTGTGGGAGAGGCTTGGCCCCCATCGGCTATG 1935
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 717 GGG 658
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1936 AGGGCGGAAGTGGGGCGGGTTGGGAGCCTCCGTGGCCCTGGCCCCCATCTCCAGTCCCC 1995
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 657 GGG 598
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1996 AGTCTGGTCCGACAGATAGGGCGAGGTGTGCTGTCTTTCAGAAGACTTCCGGCGCA 2055
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 597 GGGGGTGG 538
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2056 CATATGCATCCAGCATCTCTAAGTGGGTGGTCAAGCCACAGACTGCAGCAGAGACGCA 2115
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 537 GGGGGGGGGGAGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 478
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2116 GAAGCAAGCCAGGGACAGAGGAGAGGCCGGGAGCAGCACTTACAGCAGCAGGCAAG 2175
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 477 GGGGGGGGGTGTGG 418
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2176 GTGCGCTGGGCCCGGGCAGGCGAGTGCACGGCCACGAGAGCTCGGGG 2222
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 417 GGGGGCGGGCGGG 371

RESULT 14
BC028169

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2005, 20:40:04 ; Search time 409.498 Seconds
(without alignments)
8954.609 Million cell updates/sec

Title: US-10-615-659-27

Perfect score: 2241

Sequence: 1 tgcctgcagcagccggc.....gcctccacagggccctccc 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	69.8	3.1	7218	1	US-08-232-463-14
C 2	54.4	2.4	670689	4	US-09-949-016-12505
C 3	54.4	2.4	670690	4	US-09-949-016-14207
C 4	52.8	2.4	152331	3	US-09-128-155-16
C 5	52.2	2.3	318	3	US-09-165-264-12
C 6	51.2	2.3	319	3	US-09-165-264-8
C 7	51.2	2.3	320	3	US-09-165-264-13
C 8	51	2.3	320	3	US-09-165-264-7
C 9	51	2.3	2561	4	US-09-616-289-48
C 10	50	2.2	320	3	US-09-165-264-14
C 11	49.6	2.2	320	3	US-09-165-264-11
C 12	49.6	2.2	1926	3	US-09-249-585A-2
C 13	49.6	2.2	1926	4	US-09-410-399-3
C 14	49.6	2.2	2580	3	US-09-050-863-2
C 15	49.6	2.2	2580	3	US-09-359-081-2
C 16	49.6	2.2	5452	2	US-09-130-114-1
C 17	49.6	2.2	8705	4	US-09-647-344A-14
C 18	49.6	2.2	9600	3	US-09-910-647-1
C 19	49.6	2.2	9600	3	US-09-620-925-1
C 20	49.6	2.2	10596	1	US-07-884-811-15
C 21	49.6	2.2	10596	1	US-07-885-971-15
C 22	49.6	2.2	10596	1	US-08-087-783A-15
C 23	49.6	2.2	10596	1	US-08-194-088B-15
C 24	49.6	2.2	10596	1	US-08-194-087-15
C 25	49.6	2.2	10596	5	PCT-US93-04648-15
C 26	49.6	2.2	16080	4	US-09-724-566A-48
C 27	49.6	2.2	16080	4	US-09-471-665A-48

C 28	48.6	2.2	1247	1	US-08-278-729A-32	Sequence 32, Appl
C 29	48.6	2.2	1247	1	US-08-155-343A-32	Sequence 32, Appl
C 30	48.6	2.2	1247	1	US-08-406-672-32	Sequence 32, Appl
C 31	48.6	2.2	1247	1	US-08-643-563A-32	Sequence 32, Appl
C 32	48.6	2.2	1247	1	US-08-643-763A-32	Sequence 32, Appl
C 33	48.6	2.2	1247	1	US-08-462-623-32	Sequence 32, Appl
C 34	48.6	2.2	1247	1	US-08-451-953A-32	Sequence 32, Appl
C 35	48.6	2.2	1247	2	US-08-445-468A-32	Sequence 32, Appl
C 36	48.6	2.2	1247	2	US-08-461-397A-32	Sequence 32, Appl
C 37	48.6	2.2	1247	2	US-08-912-088-32	Sequence 32, Appl
C 38	48.6	2.2	1247	3	US-08-278-730A-32	Sequence 32, Appl
C 39	48.6	2.2	1247	3	US-08-445-467-32	Sequence 32, Appl
C 40	48.6	2.2	1247	3	US-08-480-515A-32	Sequence 32, Appl
C 41	48.6	2.2	1247	3	US-09-170-936-32	Sequence 32, Appl
C 42	48.6	2.2	1247	3	US-08-461-113-32	Sequence 32, Appl
C 43	48.6	2.2	1247	4	US-08-456-033-32	Sequence 32, Appl
C 44	48.6	2.2	1247	4	US-08-643-321-30	Sequence 30, Appl
C 45	48.6	2.2	1247	4	US-08-260-675-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match

3.1%; Score 69.8; DB 1; Length 7218;

[illegible]

```

RESULT 2
US-09-949-016-12505
; Sequence 12505, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12505
; LENGTH: 670689
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(670689)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12505

```

	Query Match	2.4%	Score 54.4;	DB 4;	Length 670689;
	Best Local Similarity	53.6%;	Pred. NO. 0.043;		
	Matches 135;	Conservative 0;	Mismatches 116;	Indels 1;	Gaps 1;
Qy	1721	CAGAGCCAGGGGCCCCGGGAGGAGGGCGGGCAGTCAAGAGTGTGTGTTCTGTGCTA	1780		
Db	40160	CCGCGGTTCGGGGGCCCGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	40219		
Qy	1781	GGCTTGTGGACAGCGCGCGGAGCAGCTGGGGCAAGCCGGGTCCTGTGTCGAGAGCCGAG	1840		
Db	40220	CGGCGTGGGCGCTGCGGGGGGGGGCGGGGGCGGGGGTGTGGGGGT-GCTGGCCCGGG	40278		

[illegible]

[illegible]

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 23, 2005, 23:48:14 ; Search time 1728.82 Seconds
(without alignment)
8666.982 Million cell updates/sec

Title: US-10-615-659-27
Perfect score: 2241
Sequence: 1 tgcctgcagcagccggc.....gcctccacagggccctccc 2241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7400704 seqs, 3343079526 residues
Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
 - 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
 - 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
 - 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
 - 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
 - 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
 - 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
 - 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
 - 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
 - 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
 - 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
 - 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
 - 22: /cgn2_6/ptodata/2/pubpna/US10J_NEW_PUB.seq.*
 - 23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
 - 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
 - 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
 - 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2241	100.0	2241	19	US-10-615-659-27	Sequence 27, Appl
2	2241	100.0	2241	19	US-10-635-977-27	Sequence 27, Appl
c 3	2241	100.0	101270	20	US-10-723-860-631	Sequence 631, App
4	1795.8	80.1	3465	19	US-10-615-659-11	Sequence 11, Appl
5	1795.8	80.1	3465	19	US-10-635-977-11	Sequence 11, Appl
6	1742.8	77.8	3554	19	US-10-615-659-12	Sequence 12, Appl
7	1742.8	77.8	3554	19	US-10-635-977-12	Sequence 12, Appl

8	272.4	12.2	1838	19	US-10-615-659-1	Sequence 1, Appli
9	272.4	12.2	1838	19	US-10-635-977-1	Sequence 1, Appli
10	272.4	12.2	1939	19	US-10-615-659-9	Sequence 9, Appli
11	272.4	12.2	1939	19	US-10-635-977-9	Sequence 9, Appli
12	138.4	6.2	1859	19	US-10-615-659-10	Sequence 10, Appl
13	138.4	6.2	1859	19	US-10-635-977-10	Sequence 10, Appl
14	69.2	3.1	452	9	US-09-728-446-1073	Sequence 1073, Ap
15	67.8	3.0	1042	22	US-10-450-763-8232	Sequence 8232, Ap
16	67.8	3.0	2250	22	US-10-450-763-5187	Sequence 5187, Ap
c 17	65	2.9	778	18	US-10-424-599-54839	Sequence 54839, A
c 18	63.2	2.8	1092	19	US-10-767-701-9739	Sequence 9739, Ap
19	62.6	2.8	740	20	US-10-425-115-131885	Sequence 131885,
20	60.8	2.7	5452	15	US-10-017-161-1481	Sequence 1481, Ap
21	60.8	2.7	5452	17	US-10-292-798-1189	Sequence 1189, Ap
c 22	60	2.7	822	20	US-10-425-115-102047	Sequence 102047, A
c 23	59.6	2.7	630	19	US-10-437-963-18774	Sequence 18774, A
c 24	59.4	2.7	891	20	US-10-425-115-146045	Sequence 146045,
c 25	58.8	2.6	766	20	US-10-425-115-82176	Sequence 82176, A
c 26	58.2	2.6	629	19	US-10-021-323-9375	Sequence 9375, Ap
c 27	57.4	2.6	588	19	US-10-021-323-9377	Sequence 9377, Ap
c 28	57.4	2.6	925	19	US-10-437-963-44536	Sequence 44536, A
c 29	57	2.5	1327	17	US-10-398-221-1265	Sequence 1265, Ap
c 30	56.8	2.5	700	20	US-10-425-115-14077	Sequence 14077, A
c 31	56.8	2.5	3163	15	US-10-017-161-1857	Sequence 1857, Ap
c 32	56.8	2.5	3163	17	US-10-292-798-1513	Sequence 1513, Ap
c 33	56.4	2.5	673	20	US-10-425-115-11145	Sequence 11145, A
c 34	56	2.5	12733	14	US-10-032-393-47	Sequence 47, Appl
c 35	56	2.5	12739	14	US-10-032-393-8	Sequence 8, Appli
c 36	55.4	2.5	959	19	US-10-437-963-48694	Sequence 48694, A
c 37	55.4	2.5	985	20	US-10-425-115-91623	Sequence 91623, A
c 38	55	2.5	3133	15	US-10-017-161-1483	Sequence 1483, Ap
c 39	55	2.5	3133	17	US-10-292-798-1191	Sequence 1191, Ap
c 40	54.8	2.4	497	18	US-10-424-599-133788	Sequence 133788,
c 41	54.6	2.4	514	18	US-10-424-599-104895	Sequence 104895,
c 42	54.4	2.4	815	18	US-10-424-599-20495	Sequence 20495, A
c 43	54.2	2.4	931	14	US-10-198-846-8585	Sequence 8585, Ap
c 44	54	2.4	771	18	US-10-424-599-107958	Sequence 107958, A
c 45	53.8	2.4	600	20	US-10-363-345A-39087	Sequence 39087, A

ALIGNMENTS

RESULT 1
US-10-615-659-27
; Sequence 27, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-27

Query Match	100.0%	Score 2241;	DB 19;	Length 2241;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2241;	Conservative	0;	Mismatches	0;
			Indels	0;
Gaps	0;			
Qy	1	TGCTGCAGCAGAGCCCGCTGGGATCTCAGCTGCAGCCATTTCCTCTGATGTT	60	
Db	1	TGCTGCAGCAGAGCCCGCTGGGATCTCAGCTGCAGCCATTTCCTCTGATGTT	60	
Qy	61	GCCTCACTCCATCCATCCTGGCTGGTGTGAACCTTGATCTGGGCATCAATAAGACT	120	

RESULT 2

US-10-635-977-27
; Sequence 27, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-27

Query Match	100.0%;	Score 2241;	DB 19;	Length 2241;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2241;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	TGCTGACGAGAGCCGGCTGGGATCCTCAGCTGCGACGCGCATTTGTCTCTGATGTT	60		
DB 1	TGCTGACGAGAGCCGGCTGGGATCCTCAGCTGCGACGCGCATTTGTCTCTGATGTT	60		
QY 61	GCCTCACTCCATCCATCTCTGCTGGTGTGAACCTTGAATGCTGGGCAATCAATAAGACT	120		
DB 61	GCCTCACTCCATCCATCTCTGCTGGTGTGAACCTTGAATGCTGGGCAATCAATAAGACT	120		
QY 121	TTTTTCTTGCCAGTCTAGTGTGTTAAATAATGAGATGCGGTACCTCTCTGGACCATCA	180		
DB 121	TTTTTCTTGCCAGTCTAGTGTGTTAAATAATGAGATGCGGTACCTCTCTGGACCATCA	180		
QY 181	GAGGACATCATTTGACTATCAGCTGACCTTACGACAGATGCTGAACCACTACGCAAA	240		
DB 181	GAGGACATCATTTGACTATCAGCTGACCTTACGACAGATGCTGAACCACTACGCAAA	240		
QY 241	GACAGCTCTTTACCA CAAAGGTGAGCGCGCGCTTGAGCTGCGGGAGCGGCA	300		
DB 241	GACAGCTCTTTACCA CAAAGGTGAGCGCGCGCTTGAGCTGCGGGAGCGGCA	300		
QY 301	GGTGGTGGGAGGAGCTTTCAGCCATTAAGGACCTCGTGGCTGCTCAGTGGCCAC	360		
DB 301	GGTGGTGGGAGGAGCTTTCAGCCATTAAGGACCTCGTGGCTGCTCAGTGGCCAC	360		
QY 361	CAGTCACCTCTCTGGGTGCATCCAGGACTCTCTGGTCTAAGGCGGTGCGGCAATCCTG	420		
DB 361	CAGTCACCTCTCTGGGTGCATCCAGGACTCTCTGGTCTAAGGCGGTGCGGCAATCCTG	420		
QY 421	GTGCCACCCCA CCCCAGCACCCCTGTGCCCTTTGCTCTGTGCTCTGGGTGAATCCGG	480		
DB 421	GTGCCACCCCA CCCCAGCACCCCTGTGCCCTTTGCTCTGTGCTCTGGGTGAATCCGG	480		
QY 481	GGCCCCAGAGTCTCCTCTAGGGCACAGAGGCCAAAGATGGGGCTGACTGGGGGCTGC	540		
DB 481	GGCCCCAGAGTCTCCTCTAGGGCACAGAGGCCAAAGATGGGGCTGACTGGGGGCTGC	540		
QY 541	CACGGGGCTTTGGGTGCTGAGGGGCTGTGGGACCCCAAGGGAGAGGTGCGGCTCCCC	600		
DB 541	CACGGGGCTTTGGGTGCTGAGGGGCTGTGGGACCCCAAGGGAGAGGTGCGGCTCCCC	600		
QY 601	CAGCACCGGGCACTCCCCCGGCTCCCGGCTCCAGGCCCTGCAAAAGAGCTTGGCAC	660		
DB 601	CAGCACCGGGCACTCCCCCGGCTCCCGGCTCCAGGCCCTGCAAAAGAGCTTGGCAC	660		

QY 661	ACCTTTCCACCAAGGCCAGGGCTGAGGCTGCCCCAGGACGACAGGCTGTGGGACCCCTG	720		
DB 661	ACCTTTCCACCAAGGCCAGGGCTGAGGCTGCCCCAGGACGACAGGCTGTGGGACCCCTG	720		
QY 721	CTGAGGAGGGGTCCCGAAGAGAGGGGCTTCCCTGGCAGAGAGTCCCTCTCAGCAGGCCA	780		
DB 721	CTGAGGAGGGGTCCCGAAGAGAGGGGCTTCCCTGGCAGAGAGTCCCTCTCAGCAGGCCA	780		
QY 781	GGTGGGCTGCTCTCAGCAGTGTGGGCGGAGGTGCAAGCAAGGTTCCCTTCGCGACC	840		
DB 781	GGTGGGCTGCTCTCAGCAGTGTGGGCGGAGGTGCAAGCAAGGTTCCCTTCGCGACC	840		
QY 841	TAAATACCCAGGGTCAGGCCAGCGCTGTCTAGTGGCAACATGCGCCCTTCAAGA	900		
DB 841	TAAATACCCAGGGTCAGGCCAGCGCTGTCTAGTGGCAACATGCGCCCTTCAAGA	900		
QY 901	CCCCGTGTGACAGCAGTCTCAGCGTGGCTGTGTTTTCTGGGGGCGCAGCATTCTTTTC	960		
DB 901	CCCCGTGTGACAGCAGTCTCAGCGTGGCTGTGTTTTCTGGGGGCGCAGCATTCTTTTC	960		
QY 961	TTAACAGGGTGACAAAC CAGACATCGGGGTGGGGACTTTCACATATGCTCTCTAATG	1020		
DB 961	TTAACAGGGTGACAAAC CAGACATCGGGGTGGGGACTTTCACATATGCTCTCTAATG	1020		
QY 1021	GGCGTGTCTCTCGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG	1080		
DB 1021	GGCGTGTCTCTCGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG	1080		
QY 1081	CAGCGCT	1140		
DB 1081	CAGCGCT	1140		
QY 1141	CCTGAGGGAGCGCCAGTGGCCACCGCTCTGGTGGCTCTCAGCTGAGCTGCTCTCTCTCT	1200		
DB 1141	CCTGAGGGAGCGCCAGTGGCCACCGCTCTGGTGGCTCTCAGCTGAGCTGCTCTCTCTCT	1200		
QY 1201	CTGGGCTCTGACGCGCTCTCAGCTGCTGAGGATGCTGAGCTGCTGAGCTGCTGAGCTG	1260		
DB 1201	CTGGGCTCTGACGCGCTCTCAGCTGCTGAGGATGCTGAGCTGCTGAGCTGCTGAGCTG	1260		
QY 1261	ATGCTCTCCACCGGTGGAGGGGTGTTCTGGGGCTCTGGGGACCGTATATCTGGGGAGG	1320		
DB 1261	ATGCTCTCCACCGGTGGAGGGGTGTTCTGGGGCTCTGGGGACCGTATATCTGGGGAGG	1320		
QY 1321	GGGAGGATCCGCTCTAGGGT	1380		
DB 1321	GGGAGGATCCGCTCTAGGGT	1380		
QY 1381	ACGGCCGTCTCTCGCTTGTGAGCGAGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	1440		
DB 1381	ACGGCCGTCTCTCGCTTGTGAGCGAGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	1440		
QY 1441	CACGCCATCTCTGAGATCGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1500		
DB 1441	CACGCCATCTCTGAGATCGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1500		
QY 1501	CCCGCCCAACCCGACTCTCTTTCCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1560		
DB 1501	CCCGCCCAACCCGACTCTCTTTCCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1560		
QY 1561	GCAGAGTCTCTGGGTAAAGTGAAGAGAGCGGAGAGGGGCCCCAGTGTCTGACAGAGAGG	1620		
DB 1561	GCAGAGTCTCTGGGTAAAGTGAAGAGAGCGGAGAGGGGCCCCAGTGTCTGACAGAGAGG	1620		
QY 1621	CTTCTAGAAAGATCCCTCTGGT	1680		
DB 1621	CTTCTAGAAAGATCCCTCTGGT	1680		
QY 1681	GACCGGGAGAGGCTCTCCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1740		
DB 1681	GACCGGGAGAGGCTCTCCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1740		
QY 1741	AGGAGGGGGGAGTCAAGAGT	1800		

Db	1741	AGGAGGGCGGGCAGTCAAGAGAGTGTGTGTTCTGTGCTAGGCTTGTGGACAGCGGGCG	1800
Qy	1801	AGCAGCTGGGGCAAGGCCGGCTCCGTGGTCSGAGGGCCGAGGGGTGCAGCTGGACGGCCGC	1860
Db	1801	AGCAGCTGGGGCAAGGCCGGCTCCGTGGTCSGAGGGCCGAGGGGTGCAGCTGGACGGCCGC	1860
Qy	1861	AGTCACAGACACACTG CAGGCGAAGAGGCGCAGG CAGAC TCGGGGTGTGGAGCAGGCTGG	1920
Db	1861	AGTCACAGACACACTG CAGGCGAAGAGGCGCAGG CAGAC TCGGGGTGTGGAGCAGGCTGG	1920
Qy	1921	GCCCCATCGGCTATGAGGGCGGGAAGTGGGGCGGGTTGGGGAGCTCTCGTGGCCCTGGCC	1980
Db	1921	GCCCCATCGGCTATGAGGGCGGGAAGTGGGGCGGGTTGGGGAGCTCTCGTGGCCCTGGCC	1980
Qy	1981	CCATCTCCAGTCCCCAGTCCCTGCTCGGACAGATAGGCGCAGGCTGTGCTGTCTTTCAG	2040
Db	1981	CCATCTCCAGTCCCCAGTCCCTGCTCGGACAGATAGGCGCAGGCTGTGCTGTCTTTCAG	2040
Qy	2041	AAGACTTCCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGGT CAGCCAC CAGAGCT	2100
Db	2041	AAGACTTCCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGGT CAGCCAC CAGAGCT	2100
Qy	2101	GCAGCAGGAC CAGCAGAGCAAGC C CAGGAC C CAGAGGAGAGAGG CCGGAGCAGCGACC	2160
Db	2101	GCAGCAGGAC CAGCAGAGCAAGC C CAGGAC C CAGAGGAGAGAGG CCGGAGCAGCGACC	2160
Qy	2161	TGAGCAGCAGGCAAGTGGCTGGCGGGCAGGCGAGTGC CAGGCCAC CAGAGCTCGG	2220
Db	2161	TGAGCAGCAGGCAAGTGGCTGGCGGGCAGGCGAGTGC CAGGCCAC CAGAGCTCGG	2220
Qy	2221	GGCTCCACAGGGGCCCTCCC	2241
Db	2221	GGCTCCACAGGGGCCCTCCC	2241
RESULT 3			
US-10-723-860-631/c			
; Sequence 631, Application US/10723860			
; Publication No US20040253606A1			
; GENERAL INFORMATION:			
; APPLICANT: Aziz, Natasha			
; APPLICANT: Ginsburg, Wendy M.			
; APPLICANT: Zlotnik, Albert			
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &			
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators			
; FILE REFERENCE: 05882.0193.NPUS01			
; CURRENT APPLICATION NUMBER: US/10/723,860			
; CURRENT FILING DATE: 2003-11-26			
; PRIOR APPLICATION NUMBER: 60/429,739			
; PRIOR FILING DATE: 2002-11-26			
; NUMBER OF SEQ ID NOS: 8393			
; SOFTWARE: Patentin version 3.2			
; SEQ ID NO 631			
; LENGTH: 101270			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-723-860-631			
Query Match 100.0%; Score 2241; DB 20; Length 101270;			
Best Local Similarity 100.0%; Pred.No. 0;			
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	TGCCTGCAGCAGAGCCGGCTGGATCCTCAGCTGCAGCCATTGTCTCTGTATGTTT	60
Db	60138	TGCCTGCAGCAGAGCCGGCTGGATCCTCAGCTGCAGCCATTGTCTCTGTATGTTT	60079
Qy	61	GCCTCACTCCATCCATCCCTGGTGGTGTGAACCTTGAATGCTGGGCATCAATAAGACT	120
Db	60078	GCCTCACTCCATCCATCCCTGGTGGTGTGAACCTTGAATGCTGGGCATCAATAAGACT	60019
Qy	121	TTTTTCTTGCAAGTCTAGGTTGGTAAATAATGAGATGCCGTACTCTCTCGGACCAATCAA	180

Db	60018	TTTTTCTTGCCAGTCTAGGTTGGTAAATAAGATGCGGTACTCTCTCTGGAACAATCA	59955
Qy	181	GAGGACATCATTTGACTATCACAGCTGACCTTACGACCAGATGCTGAACCACTACGAAA	240
Db	59958	GAGGACATCATTTGACTATCACAGCTGACCTTACGACCAGATGCTGAACCACTACGAAA	59899
Qy	241	GACAGCTCTTTCAACCAAGGTAGCCGCGCTTGAGCGTGGCGGAGCCGGCA	300
Db	59898	GACAGCTCTTTCAACCAAGGTAGCCGCGCTTGAGCGTGGCGGAGCCGGCA	59839
Qy	301	GGGTGGTGGGAGGAGCTTCTGACGCCATTAGGGACCTCGGTGGCTGGTCACTGGCCAC	360
Db	59838	GGGTGGTGGGAGGAGCTTCTGACGCCATTAGGGACCTCGGTGGCTGGTCACTGGCCAC	59779
Qy	361	CAGTCACCTCTCTGGGTGCATCCAGGACCTCTGGTCTAAGGCCGTGGGCAGAACTCATCG	420
Db	59778	CAGTCACCTCTCTGGGTGCATCCAGGACCTCTGGTCTAAGGCCGTGGGCAGAACTCATCG	59719
Qy	421	GTGCCACGCCACCCCGACGACCCCTGTGTCCTTTGTCTCTGTGTCTCTGGGTGAATCCGG	480
Db	59718	GTGCCACGCCACCCCGACGACCCCTGTGTCCTTTGTCTCTGTGTCTCTGGGTGAATCCGG	59659
Qy	481	GGCCCCAGAGCTCCCTCTCAGGGCAGAGCCCAAAGATGGGCTCACTGGGGGGTGC	540
Db	59658	GGCCCCAGAGCTCCCTCTCAGGGCAGAGGGCCAAAGATGGGCTCACTGGGGGGTGC	59599
Qy	541	CACCGGCTTTGGGTGCTCAGAGGGGCTGTGGACCCAGGGGAAAGAGTGCCTGCCCC	600
Db	59598	CACCGGCTTTGGGTGCTCAGAGGGGCTGTGGACCCAGGGGAAAGAGTGCCTGCCCC	59539
Qy	601	CAGCACCGGGCACTCCCCCGCTCCCGGGCTCCAGGCCCTGCAAAAGCAGCTTGCAC	660
Db	59538	CAGCACCGGGCACTCCCCCGCTCCCGGGCTCCAGGCCCTGCAAAAGCAGCTTGCAC	59479
Qy	661	ACCTTCCACCAAGGCCACAGGCTGAGGCTGCCCCAGGACGACGGTGTGGGACCGCTG	720
Db	59478	ACCTTCCACCAAGGCCACAGGCTGAGGCTGCCCCAGGACGACGGTGTGGGACCGCTG	59419
Qy	721	CTGAGGAGGGGTCCCGGAAGAGGGGCTTCCCTGTGCAACAGAGTCCCTCTCAGCAGGC	780
Db	59418	CTGAGGAGGGGTCCCGGAAGAGGGGCTTCCCTGTGCAACAGAGTCCCTCTCAGCAGGC	59359
Qy	781	GGTCCGCTGCCTCAGCAACAGTGTGGGGCGAGGTGACGACAAGGTTCCCTTCCGCAAC	840
Db	59358	GGTCCGCTGCCTCAGCAACAGTGTGGGGCGAGGTGACGACAAGGTTCCCTTCCGCAAC	59299
Qy	841	TAATACCCAGGTCAGGCCAGCCAGCGCTGCTAGTGGCAACATGCCCTTTCAAGA	900
Db	59298	TAATACCCAGGTCAGGCCAGCCAGCGCTGCTAGTGGCAACATGCCCTTTCAAGA	59239
Qy	901	CCCGGTGTGACAGCCAGTCAGCGTCGCTGTGTTTCTGGGGGCCAGCACTTTTTC	960
Db	59238	CCCGGTGTGACAGCCAGTCAGCGTCGCTGTGTTTCTGGGGGCCAGCACTTTTTC	59179
Qy	961	TTAACAGGGTGACAAACAGACATCGGGTGCGGGACTTACGATATGCCCTCTTAATG	1020
Db	59178	TTAACAGGGTGACAAACAGACATCGGGTGCGGGACTTACGATATGCCCTCTTAATG	59119
Qy	1021	CGCGGTGCTACTCGGTGTGGTGGCAGTGTGGGCATGTGGCTCGGTCGGGCCCTG	1080
Db	59118	CGCGGTGCTACTCGGTGTGGTGGCAGTGTGGGCATGTGGCTCGGTCGGGCCCTG	59059
Qy	1081	CAGCGCTCTCTCTCTCTGGGTCTGAGGTGGGACACCTGTGGCCGTTGGGG	1140
Db	59058	CAGCGCTCTCTCTCTCTGGGTCTGAGGTGGGACACCTGTGGCCGTTGGGG	58999
Qy	1141	CCTGGAGGGAGCCCCAGTCCCAACCGCCCTCGGTGCCCTCAGCTGACGCTCCCATTC	1200
Db	58998	CCTGGAGGGAGCCCCAGTCCCAACCGCCCTCGGTGCCCTCAGCTGACGCTCCCATTC	58939
Qy	1201	CTGGGGCTGCACGCCCTTCACTGCCTCAAGGATGCGTCTGAGCCTGGCTGCACATCC	1260
Db	58938	CTGGGGCTGCACGCCCTTCACTGCCTCAAGGATGCGTCTGAGCCTGGCTGCACATCC	58879

QY 1161 CCACGCCCTCGGTGCCCTCAGCTGCAGCCTCCCATTTCCCTGGGGCTGCCACGCCCTC 1220
DB |||||
781 CCACGCCCTCGGTGCCCTCAGCTGCAGCCTCCCATTTCCCTGGGGCTGCCACGCCCTC 840
QY 1221 CACTGCTCAAGGATGCGTCTGAGCGTGGCTGCACATCCATGCTCCACCGGTGGAGG 1280
DB |||||
841 CACTGCTCAAGGATGCGTCTGAGCGTGGCTGCACATCCATGCTCCACCGGTGGAGG 900
QY 1281 GGGTGTCTGGGGCCCTGGGGGACCGTATATCTGGGAGGGGAGGATCGCGCTTAGGG 1340
DB |||||
901 GGGTGTCTGGGGCCCTGGGGGACCGTATATCTGGGAGGGGAGGATCGCGCTTAGGG 960
QY 1341 GTGCTGTGTGTCCTCCCATGCCAAGGCAAGTCTGAGGTGGACGGCCCGTCTCCCTTGT 1400
DB |||||
961 GTGCTGTGTGTCCTCCCATGCCAGGCAAGTCTGAGGTGGACGGCCCGTCTCCCTTGT 1020
QY 1401 GAGCGAGGACAGCTGAGCTGAGTGGCCCTCTGCTTAAACAGCCACTGTCTCTGGAGA 1460
DB |||||
1021 GAGCGAGGACAGCTGAGCTGAGTGGCCCTCTGCTTAAACAGCCACTGTCTCTGGAGA 1080
QY 1461 TCGGCTGTGTGTAACATGCGGAGCCTGCGCTGCTAGCTCCCGCCAAACCCGACTCCT 1520
DB |||||
1081 TCGGCTGTGTGTAACATGCGGAGCCTGCGCTGCTAGCTCCCGCCAAACCCGACTCCT 1140
QY 1521 TCTTCCCACTGCTACAGCCTCTGCACCGAGAGTGAAGCAGGAGTTCCTGGGTAAGT 1580
DB |||||
1141 TCTTCCCACTGCTACAGCCTCTGCACCGAGAGTGAAGCAGGAGTTCCTGGGTAAGT 1200
QY 1581 GAGGAGCGGACAGAGGCCCCAGTGTCTGACGAGAGGCTTCTAGAAAGATCCCCCTG 1640
DB |||||
1201 GAGGAGCGGACAGAGGCCCCAGTGTCTGACGAGAGGCTTCTAGAAAGATCCCCCTG 1260
QY 1641 GTGCTGACAGACTGATGGGGCAGGGTCTGAGGATGAGACCGGGAGAGGCTCCCA 1700
DB |||||
1261 GTGCTGACAGACTGATGGGGCAGGGTCTGAGGATGAGACCGGGAGAGGCTCCCA 1320
QY 1701 TGGTCATGGTCATGCGAGTACAGAGGCGAGGGCCCCCGGAGGAGGCGCGGAGTCAAG 1760
DB |||||
1321 TGGTCATGGTCATGCGAGTACAGAGGCGAGGGCCCCCGGAGGAGGCGCGGAGTCAAG 1380
QY 1761 GAGTGTGTGGTCTGTGCTAGGCTTGGACAGCGCGCGAGAGCTGGGGCAGGCGCG 1820
DB |||||
1381 GAGTGTGTGGTCTGTGCTAGGCTTGTGGACAGCGCGCGAGAGCTGGGGCAGGCGCG 1440
QY 1821 CTCGTTGTTCGAGCGCGAGGGGTGACGTGCGACGGCGCAGTCAACAGAGACACTGCAGG 1880
DB |||||
1441 CTCGTTGTTCGAGCGCGAGGGGTGACGTGCGACGGCGCAGTCAACAGAGACACTGCAGG 1500
QY 1881 GAGAAAGGACAGGACACTGGGGGTGTGGGAGCAGGCTTGGGCCCCCATCGGCTATGAGGC 1940
DB |||||
1501 GAGAAAGGACAGGACACTGGGGGTGTGGGAGCAGGCTTGGGCCCCCATCGGCTATGAGGC 1560
QY 1941 GGGAAAGTGGGCGGGTGGGAGAGCTCGTGGCCCTTGGCCCATCTCGAGTCCCAGTCC 2000
DB |||||
1561 GGGAAAGTGGGCGGGTGGGAGAGCTCGTGGCCCTTGGCCCATCTCGAGTCCCAGTCC 1620
QY 2001 TGGCTCGGACAGATAGGCGGAGGCTGTGCTGCTTTTCAAGAGACTTCCGCGCACCATG 2060
DB |||||
1621 TGGCTCGGACAGATAGGCGGAGGCTGTGCTGCTTTTCAAGAGACTTCCGCGCACCATG 1680
QY 2061 GCATCCAGCATCCTCAAGTGGGTGGTCAAGCCAGAGCTGACGAGGAGCAGAGAAGC 2120
DB |||||
1681 GCATCCAGCATCCTCAAGTGGGTGGTCAAGCCAGAGCTGACGAGGAGCAGAGAAGC 1740
QY 2121 AAGCCAGGGACAGAGGAGGAGGCGGGAGAGCAGCAGCTGAGCAGGACAGAGTGC 2179
DB |||||
1741 AAGCCAGGGACAGAGGAGGAGGCGGGAGAGCAGCAGCTGAGCAGGACAGAGTGC 1799

RESULT 5
US-10-635-977-11
; Sequence 11, Application US/10635977

; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-11

Query Match 80.1%; Score 1795.8; DB 19; Length 3465;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1797; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 381 CCCAGGACTCTCTGTCTTAAGGCCGTGGCCAGAAATCACTCGGTGCCACCCACCCCGCAGC 440
DB |||||
1 CCCAGGACTCTCTGTCTTAAGGCCGTGGCCAGAAATCACTCGGTGCCACCCACCCCGCAGC 60
QY 441 ACCCTGTGCCCTTGTCTGTGTCTCTGGGTGAATCCGGGGCCCCAGAGTCCCTCTCT 500
DB |||||
61 ACCCTGTGCCCTTGTCTGTGTCTCTGGGTGAATCCGGGGCCCCAGAGTCCCTCTCT 120
QY 501 CAGGGCACAGAGGCCAAAGATGGGGCTGACTGGGGGTGCCACCGGGCTTTGGGTGCTGA 560
DB |||||
121 CAGGGCACAGAGGCCAAAGATGGGGCTGACTGGGGGTGCCACCGGGCTTTGGGTGCTGA 180
QY 561 GGGGGCTGTGGGACCCAGGGGGAAGAGTGCCTGCCCGCAGCACCGGGGACTCCCGG 620
DB |||||
181 GGGGGCTGTGGGACCCAGGGGGAAGAGTGCCTGCCCGCAGCACCGGGGACTCCCGG 240
QY 621 CCTCCCCGGCTCCAGCCCTGCACAAAGCAGCTTGCACACCTTCCCAAGAGGCCA 680
DB |||||
241 CCTCCCCGGCTCCAGCCCTGCACAAAGCAGCTTGCACACCTTCCCAAGAGGCCA 300
QY 681 GGGCTGAGGGCTGCCACAGGAGCGAGGGTGTGGGACCTTCTGCTGAGGAGGGGTCCCGAA 740
DB |||||
301 GGGCTGAGGGCTGCCACAGGAGCGAGGGTGTGGGACCTTCTGCTGAGGAGGGGTCCCGAA 360
QY 741 GAGGGGCTTCCCTGGGCA CAGAGTCCCTCTCAGCAGGCCAGGTGCGGCTGCCTCAGCACA 800
DB |||||
361 GAGGGGCTTCCCTGGGCA CAGAGTCCCTCTCAGCAGGCCAGGTGCGGCTGCCTCAGCACA 420
QY 801 GTGTGGGGCGAGGTGCAGGACAGGTTCCTTCCGACCTTAATACCCAGGGTCAGGCC 860
DB |||||
421 GTGTGGGGCGAGGTGCAGGACAGGTTCCTTCCGACCTTAATACCCAGGGTCAGGCC 480
QY 861 AGCGCCAGCGCTGTAGTGGCAACATGGCCCTTTCAAAGACCCCGTGTGCAGAGGCCAGTC 920
DB |||||
481 AGCGCCAGCGCTGTAGTGGCAACATGGCCCTTTCAAAGACCCCGTGTGCAGAGGCCAGTC 540
QY 921 AGCGTGCCTGTGTGTTCCTGGGGCCAGCACAATTTTTTCTTAAACAGGTGACAAACAG 980
DB |||||
541 AGCGTGCCTGTGTGTTCCTGGGGCCAGCACAATTTTTTCTTAAACAGGTGACAAACAG 600
QY 981 ACATCGGGGTGGGGGACTTCAGATATGCTCTCTAATGCGCGTGTCTACCTCGCTGT 1040
DB |||||
601 ACATCGGGGTGGGGGACTTCAGATATGCTCTCTAATGCGCGTGTCTACCTCGCTGT 660
QY 1041 GGTGGGAGTGTGGGCAATGTGGCTGGTTCCGGGCCCTTCAGCGCTCTCTCTCTCTCT 1100
DB |||||
661 GGTGGGAGTGTGGGCAATGTGGCTGGTTCCGGGCCCTTCAGCGCTCTCTCTCTCTCT 720
QY 1101 GGGGTCTGAGGTGTGGGACACCCCTGTGGCCCGGTTTGGGGGCTGGAGGGAGCCCCAGTGC 1160


```

Db 721 GGGGCTGAGGTGTGGGACACCTGTGGCCCGTTTGGGGGCTTGGAGGAGCCCCAGTGC 780
QY 1161 CCACGCCCTCGGTGCTCAGTGCAGGCTTCCCAATTCCTTGAGGCTTGCACAGCCCTC 1220
Db 781 CCACGCCCTCGGTGCTCAGTGCAGGCTTCCCAATTCCTTGAGGCTTGCACAGCCCTC 840
QY 1221 CACTGCTCAAGGATGCGTCTGAGCTGCGCTGCGCATCATCATCTCCACCGGCTGGAG 1280
Db 841 CACTGCTCAAGGATGCGTCTGAGCTGCGCTGCGCATCATCATCTCCACCGGCTGGAG 900
QY 1281 GGGTGTTCCTGGGGCCCTGGGGACCGTATATCTGGGGAGGGGAGGATCCGCCCTTAGG 1340
Db 901 GGGTGTTCCTGGGGCCCTGGGGACCGTATATCTGGGGAGGGGAGGATCCGCCCTTAGG 960
QY 1341 GTGCTGTGTGGTCCCCCATAGCCAGCAAGTCTGAGGTGGAGCGGCCGCTCTGCGCTGT 1400
Db 961 GTGCTGTGTGGTCCCCCATAGCCAGCAAGTCTGAGGTGGAGCGGCCGCTCTGCGCTGT 1020
QY 1401 GAGCGAGGCGAGAGCTGAGCTGAGCTGCCCTCTCTGCTTAACACGCCACTGTCTTGGAGA 1460
Db 1021 GAGCGAGGCGAGAGCTGAGCTGAGCTGCCCTCTCTGCTTAACACGCCACTGTCTTGGAGA 1080
QY 1461 TCGGGCTGTGCTGAACATGCGGAGCCTGCCCTGTGTACGTCCCGGCCAACCCCGACTCT 1520
Db 1081 TCGGGCTGTGCTGAACATGCGGAGCCTGCCCTGTGTACGTCCCGGCCAACCCCGACTCT 1140
QY 1521 TCTTCCCAAGCTGTACAGCCTTGCACCGAGAGTGAGCAGCAGGATTCCTGGGTAAAGT 1580
Db 1141 TCTTCCCAAGCTGTACAGCCTTGCACCGAGAGTGAGCAGCAGGATTCCTGGGTAAAGT 1200
QY 1581 GAGGAGAGGCGAGAGGCCCCAGTGTGTACGACGAGAGGCTTCTAGAAAGATCCCCCTG 1640
Db 1201 GAGGAGAGGCGAGAGGCCCCAGTGTGTACGACGAGAGGCTTCTAGAAAGATCCCCCTG 1260
QY 1641 GTGCTGAGACAGACTGATGGGCGAGGCTGTGAGGATAGAGGACCGGGGAGAGGCTCCCA 1700
Db 1261 GTGCTGAGACAGACTGATGGGCGAGGCTGTGAGGATAGAGGACCGGGGAGAGGCTCCCA 1320
QY 1701 TGGTCAATGTCATGTCAGTACAGAGCCAGGGCCCCCGGAGGGAGGGCGGCGACTCAAG 1760
Db 1321 TGGTCAATGTCATGTCAGTACAGAGCCAGGGCCCCCGGAGGGAGGGCGGCGACTCAAG 1380
QY 1761 GAGTGTGTGGTTCCTGTGCTAGGCTGTGAGCAGCGGGGCGAGCAGCTGGGGCAAGCCGG 1820
Db 1381 GAGTGTGTGGTTCCTGTGCTAGGCTTTGTGGACAGCGGCGAGCAGCTGGGGCAAGCCGG 1440
QY 1821 CTCCTGTGTGGAGCCGAGGGGTGCAGCTGGACGCGCGCAGTCAAGAGACACTGCAAG 1880
Db 1441 CTCCTGTGTGGAGCCGAGGGGTGCAGCTGGACGCGCGCAGTCAAGAGACACTGCAAG 1500
QY 1881 GAGAGGGCGAGCAGACTGGGGGTGTGGAGCAGGCTTGGGCCCCCATCGGCTATGAGGGC 1940
Db 1501 GAGAGGGCGAGCAGACTGGGGGTGTGGAGCAGGCTTGGGCCCCCATCGGCTATGAGGGC 1560
QY 1941 GGGAGTGGGGCGGTTGGGAGACCTCGTGGCCCTTGGCCCATCTCCAGTCCCGAGTCC 2000
Db 1561 GGGAGTGGGGCGGTTGGGAGACCTCGTGGCCCTTGGCCCATCTCCAGTCCCGAGTCC 1620
QY 2001 TGGCTCGGACAGATAGGGCGAGGCTGTGTGCTTTTCAAGAGACTTCCGGGCGACCATG 2060
Db 1621 TGGCTCGGACAGATAGGGCGAGGCTGTGTGCTTTTCAAGAGACTTCCGGGCGACCATG 1680
QY 2061 GCATCCAGCATCCTCAAGTGGTGGTGCAGCCACAGAGCTGCAGCAGGAGCAGCAGAGC 2120
Db 1681 GCATCCAGCATCCTCAAGTGGTGGTGCAGCCACAGAGCTGCAGCAGGAGCAGCAGAGC 1740
QY 2121 AAGCCAGGACACAGAGGAGAGGCGGGAGCGAGCAGCAGCCTGAGCAGCAGGCAAGGTGC 2179
Db 1741 AAGCCAGGACACAGAGGAGAGGCGGGAGCGAGCAGCAGCCTGAGCAGCAGGCAAGTGC 1799
```

RESULT 6

```

US-10-615-659-12
; Sequence 12, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-12
```

```

Query Match 77.8%; Score 1742.8; DB 19; Length 3554;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1784; Conservative 0; Mismatches 2; Indels 30; Gaps 1;

QY 394 GTCTAAGGCCGTGGCCAGCAATCACTCGGTGCCACCCACCCAGCAGCCCTGTGCCCT 453
Db 1 GTCTAAGGCCGTGGCCAGCAATCACTCGGTGCCACCCACCCAGCAGCCCTGTGCCCT 60
QY 454 TTGCTCTGTGCTCTCTGGGTGAATCCGGGGCCCCAGAAAGCTCCCTCTCAGGGCAGAGG 513
Db 61 TTGCTCTGTGCTCTCTGGGTGAATCCGGGGCCCCAGAAAGCTCCCTCTCAGGGCAGAGG 120
QY 514 CCAAGATGGGGCTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGAGGGGCTGTGGGA 573
Db 121 CCAAGATGGGGCTGACTGGGGGCTGCCACCGGGCTTTGGGTGCTGAGGGGCTGTGGGA 180
QY 574 CCCAGGGGAGAGGTGCGCTCCCGCAGACCGGGCACTCCCGCGCTCCCGGCT 633
Db 181 CCCAGGGGAGAGGTGCGCTCCCGCAGACCGGGCACTCCCGCGCTCCCGGCT 240
QY 634 CCAGCCCTGCACAAAGCAGCTTGCCACACCTTCCACAAAGGCCAGGGCTGAGGCCCTG 693
Db 241 CCAGCCCTGCACAAAGCAGCTTGCCACACCTTCCACAAAGGCCAGGGCTGAGGCCCTG 300
QY 694 CCCAGGACGCGAGGTGGGGAACCTGTGTAGAGGAGGGTCCCGAAGAGGGGCTTCCT 753
Db 301 CCCAGGACGCGAGGTGGGGAACCTGTGTAGAGGAGGGTCCCGAAGAGGGGCTTCCT 360
QY 754 GGCACAGAGTCCCTCTCAGCAGGCGGCTGGGCTGCTCAGCAGACTGTGGGGGGAG 813
Db 361 GGCACAGAGTCCCTCTCAGCAGGCGGCTGGGCTGCTCAGCAGACTGTGGGGGGAG 420
QY 814 GTGAGGACAAAGTTCCCTTCCGCACTTAATACCCAGGGTCAAGGCCAGCGCGCTG 873
Db 421 GTGAGGACAAAGTTCCCTTCCGCACTTAATACCCAGGGTCAAGGCCAGCGCGCTG 480
QY 874 CTAGTGGCAAATGCCCCCTTCAAAGACCCCGTGTGAGAGCCAGTCAAGTGCAGTGTG 933
Db 481 CTAGTGGCAAATGCCCCCTTCAAAGACCCCGTGTGAGAGCCAGTCAAGTGCAGTGTG 540
QY 934 TTTTCTGGGGCGCAGCAGCTTTTCTTAACAGGGTGACAAACAGACATCGGGGTGG 993
Db 541 TTTTCTGGGGCGCAGCAGCTTTTCTTAACAGGGTGACAAACAGACATCGGGGTGG 600
QY 994 GGGACTTCACGATATGCTCTCTAATGGCGGTGCTACTCGCGTGTGGTGGGCAAGTGT 1053
Db 601 GGGACTTCACGATATGCTCTCTAATGGCGGTGCTACTCGCGTGTGGTGGGCAAGTGT 660
QY 1054 GGGCATGTGGTGTGCTTCCGGGCCCTGCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1113
Db 661 GGGCATGTGGTGTGCTTCCGGGCCCTGCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 720
QY 1114 TGGGACACCCCTGTGGCCGCTTTTGGGGGCTGTGGAGGAGCCCGCAGTGTGCCACCGCCCTCG 1173
```


Db 721 TGGGACACCTGTGGCCCGCTTGGGGGCTCGAGGGAGCCCGCAGTGGCCACCGCCCTCGG 780
QY 1174 TGCCCTCAGCTCAGCCTCCCATTCCTCGGGCTGCCACGCGCTCCACTCGCTCAAGG 1233
Db 781 TGCCCTCAGCTCAGCCTCCCATTCCTCGGGCTGCCACGCGCTCCACTCGCTCAAGG 840
QY 1234 ATGCGTCTGAGCCTGGCCTGCACATCCATGTCTCCACCGGCTGGAGGGGGTTCCTGGG 1293
Db 841 ATGGTCTGAGCCTGGCCTGCACATCCATGTCTCCACCGGCTGGAGGGGGTTCCTGGG 900
QY 1294 CCCTGGGGGACCGTATATCTGGGAGGGGGAGGATCCGCCCTAGGGGTGCTGTGTGGT 1353
Db 901 CCCTGGGGGACCGTATATCTGGGAGGGGGAGGATCCGCCCTAGGGGTGCTGTGTGGT 960
QY 1354 CCCCATAGCCAGGCAAGTCTCAGGTGGACGCGCGCTCTCGCTTGTGAGGAGGCGAG 1413
Db 961 CCCCATAGCCAGGCAAGTCTCAGGTGGACGCGCGCTCTCGCTTGTGAGGAGGCGAG 1020
QY 1414 CTGAGCTGAGCTGCCCTCTCTGCCCTAAACAACGCCACTGTCTCTGGAGATCGGGCTGTGCGT 1473
Db 1021 CTGAGCTGAGCTGCCCTCTCTGCCCTAAACAACGCCACTGTCTCTGGAGATCGGGCTGTGCGT 1080
QY 1474 GAAATGCGGAGCCTGCCCTCGTGTAGTCCCGGCCAAACCCGACTCTCTTCCACGCTG 1533
Db 1081 GAAATGCGGAGCCTGCCCTCGTGTAGTCCCGGCCAAACCCGACTCTCTTCCACGCTG 1140
QY 1534 CTACAGCCTCTGCACCGAGAGTGAGCAGCAGAGTTCCTGGTGAAGTGAGGAGAGCGGAG 1593
Db 1141 CTACAGCCTCTGCACCGAGAGTGAGCAGCAGAGTTCCTGGTGAAGTGAGGAGAGCGGAG 1200
QY 1594 AGGGCCCCAGTCTGTACAGCAGAGAGGCTTCTAGAAAGATCCCCCTGTGTCTGAGACAGA 1653
Db 1201 AGGGCCCCAGTCTGTACAGCAGAGAGGCTTCTAGAAAGATCCCCCTGTGTCTGAGACAGA 1260
QY 1654 CTGATGGGGCAGGGTCTGAGGATAGAGAACCGGGGAGAGGCTCCCATGTGTCATGTGTCAT 1713
Db 1261 CTGATGGGGCAGGGTCTGAGGATAGAGAACCGGGGAGAGGCTCCCATGTGTCATGTGTCAT 1320
QY 1714 GGCAGTACAGAGGCGAGGGGCCCGGAGGGAGCGGCGGAGTCAAGGAGTGTGTGTTTC 1773
Db 1321 GGCAGTACAGAGGCGAGGGGCCCGGAGGGAGGGCGGCGGAGTCAAGGAGTGTGTGTTTC 1380
QY 1774 TGTGTAGGCTTGTGGACAGCGCGGCGAGCAGCTGGGGCAAGGCGGCTCCGTGGTCGGA 1833
Db 1381 TGTGTAGGCTTGTGGACAGCGCGGCGAGCAGCTGGGGCAAGGCGGCTCCGTGGTCGGA 1440
QY 1834 GGCAGAGGGTGCAGCTGGACGCGCGGCGAGTCAAGAGACATGACGAGGAGAGGGCAGGC 1893
Db 1441 GGCAGAGGGTGCAGCTGGACGCGCGGCGAGTCAAGAGACATGACGAGGAGAGGGCAGGC 1500
QY 1894 AGACTGGGGGTGGGAGCAGGCTGGGCGCCCATCGGCTATGAGGGCGGGAAGTGGGGCG 1953
Db 1501 AGACTGGGGGTGGGAGCAGGCTGGGCGCCCATCGGCTATGAGGGCGGGAAGTGGGGCG 1560
QY 1954 GGTGGGGAGCCTCGTGGCCCTGGCCCCATCTCCAGTCCCCAGTCTGTGGTTCGGAACAGA 2013
Db 1561 GGTGGGGAGCCTCGTGGCCCTGGCCCCATCTCCAGTCCCCAGTCTGTGGTTCGGAACAGA 1620
QY 2014 TAGGGCAGAGGCTGTGCTGTCTTTCA-----GAAG 2043
Db 1621 TAGGGCAGAGGCTGTGCTGTCTTTTCAGAGTAGCAGCGTGGGAAGAGAGTTCCTGGAG 1680
QY 2044 ACTTCCGGCGCACCATGCGATCCAGCATCTCAAGTGGGTGTCAGCCACAGAGCTGCA 2103
Db 1681 ACTTCCGGCGCACCATGCGATCCAGCATCTCAAGTGGGTGTCAGCCACAGAGCTGCA 1740
QY 2104 GCAGAGCAGCAGAGCAAGCCAGGACACAGAGGAGGAGCGGGGAGCAGCAGCTGA 2163
Db 1741 GCAGAGCAGCAGAGCAAGCCAGGACACAGAGGAGGAGCGCGGGAGCAGCGACCTGA 1800
QY 2164 GCAGAGCAGAGGTGC 2179
|||||

Db 1801 GCACAGGCAAGATGC 1816
RESULT 7
US-10-635-977-12
; Sequence 12, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 3554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-12
Query Match 77.8%; Score 1742.8; DB 19; Length 3554;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1784; Conservative 0; Mismatches 2; Indels 30; Gaps 1;
QY 394 GTCTAAGGCGGTGGCCAGAAATCACTCGGTGCCCCACCCACCCCGACACCCCTGTGCCCT 453
Db 1 GTCTAAGGCGGTGGCCAGAAATCACTCGGTGCCCCACCCACCCCGACACCCCTGTGCCCT 60
QY 454 TTGCTCTGTCTCTCTGGTGAATCCGGGGCCCGAGAGCTCCCTCCTCAGGGCACAGG 513
Db 61 TTGCTCTGTCTCTCTGGTGAATCCGGGGCCCGAGAGCTCCCTCCTCAGGGCACAGG 120
QY 514 CCAAGATGGGGCTGACTGGGGGCTGCACCGGGCTTTGGGTGCTGAGGGGGCTGTGGGA 573
Db 121 CCAAGATGGGGCTGACTGGGGGCTGCACCGGGCTTTGGGTGCTGAGGGGGCTGTGGGA 180
QY 574 CCCAGGGGAAGAGTCCGCTCCCTCCAGAACCGGGCACTCCCGCGCTCCCGCGGCT 633
Db 181 CCCAGGGGAAGAGTCCGCTCCCTCCAGAACCGGGCACTCCCGCGCTCCCGCGGCT 240
QY 634 CCAGCCCTGCACAAAGCAGCTTGCACACCTTCCCAAGGCCAGGGCTGAGGGCTG 693
Db 241 CCAGCCCTGCACAAAGCAGCTTGCACACCTTCCCAAGGCCAGGGCTGAGGGCTG 300
QY 694 CCCAGGACGAGGGTGTGGGGAACCTGCTGAGGGAGGGGTCCCGAAGAGAGGGCTTCCCT 753
Db 301 CCCAGGACGAGGGTGTGGGGAACCTGCTGAGGGAGGGGTCCCGAAGAGAGGGGTTCCT 360
QY 754 GGCAACAGAGTTCCTTCTCAGACAGGCCAGGTGCGGCTCCCTCAGCACAGTGTGGGGCGAG 813
Db 361 GGCAACAGAGTTCCTTCTCAGACAGGCCAGGTGCGGCTCCCTCAGCACAGTGTGGGGCGAG 420
QY 814 GTGCAGACAGGTTCCTTCCGACCTAATACCCAGGGTCCAGGCCAGCGCCAGCGCTG 873
Db 421 GTGCAGACAGGTTCCTTCCGACCTAATACCCAGGGTCCAGGCCAGCGCCAGCGCTG 480
QY 874 CTAGTGGCAATATGGCCCTTCAAGAACCCCGCTGTGCAGAGCCAGTCAAGGTGCGCTGTG 933
Db 481 CTAGTGGCAATATGGCCCTTCAAGAACCCCGCTGTGCAGAGCCAGTCAAGGTGCGCTGTG 540
QY 934 TTTTCTGGGGCCAGCCACATTTTCTTAAAGGGTGAACAAACAGACATCGGGGTGCG 993
Db 541 TTTTCTGGGGCCAGCCACATTTTCTTAAAGGGTGAACAAACAGACATCGGGGTGCG 600
QY 994 GGGACTTTCAGATATGCTCTTAATGGCGGTGTACCTCGCTGCTGGGAGTGTCT 1053
Db 601 GGGACTTTCAGATATGCTCTTAATGGCGGTGTACCTCGCTGCTGGGAGTGTCT 660

QY 1054 GGGCATGTGGCTGCGTTTCGGGGCCCTGCAAGCCGCTCCTCCCTCTCTCTGGGGTCTGAGGTG 1113
DB |||||||
QY 661 GGGCATGTGGCTGCGTTTCGGGGCCCTGCAAGCCGCTCCTCCCTCTCTGGGGTCTGAGGTG 720
DB |||||||
QY 1114 TGGGACACCTGTGGCCCGCTTTGGGGGCTGAGAGGAGCCCAAGTCCACCGCCCTCGG 1173
DB |||||||
QY 721 TGGGACACCTGTGGCCCGCTTTGGGGGCTGAGAGGAGCCCAAGTCCACCGCCCTCGG 780
DB |||||||
QY 1174 TGGCCCTGAGCTGAGCTGCGCTCCCAATTCCTCTGGGGCTGCGCAAGCCCTCCACGTCCCAAGG 1233
DB |||||||
QY 781 TGGCCCTGAGCTGAGCTGCGCTCCCAATTCCTCTGGGGCTGCGCAAGCCCTCCACGTCCCAAGG 840
DB |||||||
QY 1234 ATGGCTCTGAGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGG 1293
DB |||||||
QY 841 ATGGCTCTGAGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGG 900
DB |||||||
QY 1294 CCTTGGGGGACCGTATATCTGGGGGAGGGAGGATCCGCTTGGGGTGTCTGTGGTG 1353
DB |||||||
QY 901 CCTTGGGGGACCGTATATCTGGGGGAGGGAGGATCCGCTTGGGGTGTCTGTGGTG 960
DB |||||||
QY 1354 CCCCATAGCCAGGCAAGTCTGAGGTGAGCGGCCCTCTCGCTTGTGAGCGAGGACAG 1413
DB |||||||
QY 961 CCCCATAGCCAGGCAAGTCTGAGGTGAGCGGCCCTCTCGCTTGTGAGCGAGGACAG 1020
DB |||||||
QY 1414 CTGAGCTGAGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGG 1473
DB |||||||
QY 1021 CTGAGCTGAGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGG 1080
DB |||||||
QY 1474 GAAATGCGGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1533
DB |||||||
QY 1081 GAAATGCGGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
DB |||||||
QY 1534 CTACAGCTCTGCAACCGAGAGTGAGCAGCAGGAGTTCCTGGTAAAGTGAAGGAGCGGAG 1593
DB |||||||
QY 1141 CTACAGCTCTGCAACCGAGAGTGAGCAGCAGGAGTTCCTGGTAAAGTGAAGGAGCGGAG 1200
DB |||||||
QY 1594 AGGGCCCCAGTGTCTGAGCAGAGAGGCTTCTAGAAAGATCCCTGCTGCTGAGACAGA 1653
DB |||||||
QY 1201 AGGGCCCCAGTGTCTGAGCAGAGAGGCTTCTAGAAAGATCCCTGCTGCTGAGACAGA 1260
DB |||||||
QY 1654 CTGATGGGGGAGGCTGAGAGATGAGAGCAGGGGAGAGGCTCCCATGCTCATGCTGCTCAT 1713
DB |||||||
QY 1261 CTGATGGGGGAGGCTGAGAGATGAGAGCAGGGGAGAGGCTCCCATGCTCATGCTGCTCAT 1320
DB |||||||
QY 1714 GGCAGTACAGAGGCGCAGGGGCCCCCGGAGGAGGCGCGGCAAGTCAAGGAGTGTGTGTTTC 1773
DB |||||||
QY 1321 GGCAGTACAGAGGCGCAGGGGCCCCCGGAGGAGGCGCGGCAAGTCAAGGAGTGTGTGTTTC 1380
DB |||||||
QY 1774 TGTGTAGGCTTGTGGAAGCAGCGCGGAGCAGCTGGGGGAAAGGCGCGCTCCGTGCTGGA 1833
DB |||||||
QY 1381 TGTGTAGGCTTGTGGAAGCAGCGCGGAGCAGCTGGGGGAAAGGCGCGCTCCGTGCTGGA 1440
DB |||||||
QY 1834 GSCCAGAGGGGTGAGCTGGAAGGCGCGCAGTCAAGAGACCTGAGGAGGAGGAGGCGAGC 1893
DB |||||||
QY 1441 GSCCAGAGGGGTGAGCTGGAAGGCGCGCAGTCAAGAGACCTGAGGAGGAGGAGGCGAGC 1500
DB |||||||
QY 1894 AGACTGGGGGTGTGGAGCAGGCTGCGGCCCCCATCGGCTATGAGGCGGGGAAAGTGGGGCG 1953
DB |||||||
QY 1501 AGACTGGGGGTGTGGAGCAGGCTGCGGCCCCCATCGGCTATGAGGCGGGGAAAGTGGGGCG 1560
DB |||||||
QY 1954 GGTGGGAGGCTTCCGTGGCCCTGCGCCCATCTCCAGTCCCGAGTCTGGCTCGGACAGA 2013
DB |||||||
QY 1561 GGTGGGAGGCTTCCGTGGCCCTGCGCCCATCTCCAGTCCCGAGTCTGGCTCGGACAGA 1620
DB |||||||
QY 2014 TAGGGCGAGGCTGCTGCTGCTTCA-----GAG 2043
DB |||||||
QY 1621 TAGGGCGAGGCTGCTGCTGCTTTCAGAGTACAGCGTGGGAAAGAGGTTCCTGGGAAG 1680
DB |||||||
QY 2044 ACTTCGGGCGGACCATGAGCATCCAGCATCTCAAGTGGGTGTGAGCCACAGAGCTGCA 2103
DB |||||||
QY 1681 ACTTCGGGCGGACCATGAGCATCCAGCATCTCAAGTGGGTGTGAGCCACAGAGCTGCA 1740
DB |||||||

QY 2104 GCAGGAGCAGCAGAGCAAGCCCGAGGACCAAGGAGGAGCGCGGAGCAGCGACCTGA 2163
DB |||||||
QY 1741 GCAGGAGCAGCAGAGCAAGCCCGAGGACCAAGGAGGAGCGCGGAGCAGCGACCTGA 1800
DB |||||||
QY 2164 GCAGCAGGCAAGGTGC 2179
DB |||||||
QY 1801 GCAGCAGGCAAGATGC 1816
DB |||||||
RESULT 8
US-10-615-659-1
; Sequence 1, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-615-659-1
Query Match 12.2%; Score 272.4; DB 19; Length 1838;
Best Local Similarity 99.6%; Pred. No. 9.8e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1906 TGGGAGCAGGCTGGGCCCCCATCTCCAGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1965
DB 1 TGGGAGCAGGCTGGGCCCCCATCTCCAGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60
QY 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2025
DB 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
QY 2026 GTGCTGTCTTTTCAGAGACTTCGCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG 2085
DB 121 GTGCTGTCTTTTCAGAGACTTCGCGGCGCACCATGGCATCCAGCATCTCAAGTGGGTGG 180
QY 2086 TCAGCCACCAAGCTGCGAGCAGGAGCAGAGCAAGCAAGCCAGGAGCCAGAGGAGGAGG 2145
DB 181 TCAGCCACCAAGCTGCGAGCAGGAGCAGAGCAAGCAAGCCAGGAGCCAGAGGAGGAGG 240
QY 2146 CCGGAGCAGGAGCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2179
DB 241 CCGGAGCAGGAGCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 274
RESULT 9
US-10-635-977-1
; Sequence 1, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09


```
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (153)..(1775)
US-10-635-977-1

Query Match      12.2%; Score 272.4; DB 19; Length 1838;
Best Local Similarity 99.6%; Pred. No. 9.8e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGCGGTTGGGGAGCC 1965
DB 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGCGGTTGGGGAGCC 60

QY 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGAGGCT 2025
DB 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGAGGCT 120

QY 2026 GTGCTGTCTTTTCAAGAAGACTTCCGGCGGCAACATGGCATCCAGCATCTCAAGTGGGTGG 2085
DB 121 GTGCTGTCTTTTCAAGAAGACTTCCGGCGGCAACATGGCATCCAGCATCTCAAGTGGGTGG 180

QY 2086 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGACCCAGGAGGAGGAGG 2145
DB 181 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGAGGAGGAGGAGGAGG 240

QY 2146 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 2179
DB 241 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 274

RESULT 10
US-10-615-659-9
; Sequence 9, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-9

Query Match      12.2%; Score 272.4; DB 19; Length 1939;
Best Local Similarity 99.6%; Pred. No. 9.9e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGCGGTTGGGGAGCC 1965
DB 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGCGGTTGGGGAGCC 60

QY 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGAGGCT 2025
DB 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGAGGCT 120

QY 2026 GTGCTGTCTTTTCAAGAAGACTTCCGGCGGCAACATGGCATCCAGCATCTCAAGTGGGTGG 2085
DB 121 GTGCTGTCTTTTCAAGAAGACTTCCGGCGGCAACATGGCATCCAGCATCTCAAGTGGGTGG 180

QY 2086 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGACCCAGGAGGAGGAGGAGG 2145
DB 181 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGAGGAGGAGGAGGAGGAGG 240

QY 2146 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 2179
DB 241 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 274

RESULT 11
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match      12.2%; Score 272.4; DB 19; Length 1939;
Best Local Similarity 99.6%; Pred. No. 9.9e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGCGGTTGGGGAGCC 1965
DB 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGCGGTTGGGGAGCC 60

QY 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGAGGCT 2025
DB 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGAGGCT 120

QY 2026 GTGCTGTCTTTTCAAGAAGACTTCCGGCGGCAACATGGCATCCAGCATCTCAAGTGGGTGG 2085
DB 121 GTGCTGTCTTTTCAAGAAGACTTCCGGCGGCAACATGGCATCCAGCATCTCAAGTGGGTGG 180

QY 2086 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGACCCAGGAGGAGGAGGAGG 2145
DB 181 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGAGGAGGAGGAGGAGGAGG 240

QY 2146 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 2179
DB 241 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 274

RESULT 12
US-10-615-659-10
; Sequence 10, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
```

```
QY 2086 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGACCCAGGAGGAGGAGG 2145
DB 181 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGAGGAGGAGGAGGAGGAGG 240

QY 2146 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 2179
DB 241 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 274

RESULT 11
US-10-635-977-9
; Sequence 9, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-9

Query Match      12.2%; Score 272.4; DB 19; Length 1939;
Best Local Similarity 99.6%; Pred. No. 9.9e-64;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1906 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGCGGTTGGGGAGCC 1965
DB 1 TGGGAGCAGGCTGGGCCCCCATCGCTATGAGGGCGGGAAGTGGGCGGTTGGGGAGCC 60

QY 1966 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGAGGCT 2025
DB 61 TCCGTGGCCCTGGGCCCCCATCTCCAGTCCCCAGTCTCTGGCTCGGACAGATAGGGGAGGCT 120

QY 2026 GTGCTGTCTTTTCAAGAAGACTTCCGGCGGCAACATGGCATCCAGCATCTCAAGTGGGTGG 2085
DB 121 GTGCTGTCTTTTCAAGAAGACTTCCGGCGGCAACATGGCATCCAGCATCTCAAGTGGGTGG 180

QY 2086 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGACCCAGGAGGAGGAGGAGG 2145
DB 181 TCAGCCACAGAGCTGCAGCAGGAGCAGCAGACAGCCAGGAGGAGGAGGAGGAGGAGG 240

QY 2146 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 2179
DB 241 CCGGAGCAGGACCTGAGCAGCAGGCAAGGTGC 274

RESULT 12
US-10-615-659-10
; Sequence 10, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT FILING DATE: 2003-07-09
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
```



```
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-615-659-10

Query Match
Best Local Similarity 6.2%; Score 138.4; DB 19; Length 1859;
Matches 139; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2040 GAAGACTTCGGCGCACCACCATGCGATCCAGCATCTCAAGTGGTGGTTCAGCCACGAGC 2099
Db 29 GAAGACTTCGGCGCACCACCATGCGATCCAGCATCTCAAGTGGTGGTTCAGCCACGAGC 88
QY 2100 TCCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2159
Db 89 TCCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 148
QY 2160 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2179
Db 149 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 168

RESULT 13
US-10-635-977-10
; Sequence 10, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 10
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-635-977-10

Query Match
Best Local Similarity 6.2%; Score 138.4; DB 19; Length 1859;
Matches 139; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2040 GAAGACTTCGGCGCACCACCATGCGATCCAGCATCTCAAGTGGTGGTTCAGCCACGAGC 2099
Db 29 GAAGACTTCGGCGCACCACCATGCGATCCAGCATCTCAAGTGGTGGTTCAGCCACGAGC 88
QY 2100 TCCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2159
Db 89 TCCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 148
QY 2160 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2179
Db 149 CTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 168

RESULT 14
US-09-728-446-1073
; Sequence 1073, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020081668A1el Murine Polynucleotide Sequences
; FILE REFERENCE: LEX-0101-1073
; OTHER INFORMATION: and Mutant Cells and Mutant Animals Defined Thereby
```

```
; CURRENT APPLICATION NUMBER: US/09/728,446
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,270
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 1461
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1073
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Mus musculus
; NAME/KEY: misc feature
; LOCATION: (1)-(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-446-1073

Query Match
Best Local Similarity 3.1%; Score 69.2; DB 9; Length 452;
Matches 95; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 132 AGTCTAGTTGGTAAATAATGAGATGCGGTACCTCTCTGACCATCAAGAGGGACATCA 191
Db 6 AGTCCCGCATGGTTCGAAATGAGACCCCTACCTCATCTGGACCACTCGGCGGGATGTGC 65
QY 192 TTGACTATCAGAGCTGACCTACGACGAGATGCTGAACCACTACGCAAGAGACGCTCT 251
Db 66 TGGATTGTGCTTCTCTTCCAAAGATCAGATGATAAACCATTATGCCCGTGCAGGCTCT 125
QY 252 TCACCACCAAGGTGAGCC 269
Db 126 TCACCACCAAGGTGAGCC 143

RESULT 15
US-10-450-763-8232
; Sequence 8232, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 8232
; LENGTH: 1042
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: SIMILAR
; LOCATION: (346)-(35)
; OTHER INFORMATION: 35% homologous to Streptomyces ansochromogenes SanE, accession
; OTHER INFORMATION: number AF228524, Smith-Waterman Score=100.
US-10-450-763-8232

Query Match
Best Local Similarity 3.0%; Score 67.8; DB 22; Length 1042;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 96 TGAATGCTGGGCATCAATAAGACTTTTCTTCCAGTCTAGGTGGTAAATAATGAGA 155
Db 707 TGAATTTGATGACCTAGATGGAACACATGCTCTGATGTCCCGCATGTGTCAGAAATGAGA 766
QY 156 TGGCGTACCTCTCTGGACCATCAAGAGGGACATCATGACTATCAGAGCTGACCTACG 215
Db 767 TCCCTACTTTCATCTGGACCACTCGGCGGGATGTGCTCGACTGTCTTCTCTCCAAAGG 826
```


Qy 216 ACCAGATGCTGAACCACTACGCAAGACAGCCTCCTTCACCAAG 262
Db 827 ATCAGATGATGAACCACTACGCGGGCTGGCTCCTTTACCAAG 873

Search completed: September 24, 2005, 16:48:46
Job time : 1731.82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2005, 09:00:19 ; Search time 43 Seconds
(without alignments)
1210.541 Million cell updates/sec

Title: US-10-615-659-2
Perfect score: 2854
Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTAEGALRPPPGKGS f41
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	956.5	33.5	352	2 T12515	hypothetical prote
2	259	9.1	379	2 A45443	tubulin-tyrosine 1
3	252.5	8.8	640	2 E88575	protein ZK1128.6
4	252.5	8.8	680	2 T27699	hypothetical prote
5	209.5	7.3	1198	2 T20262	hypothetical prote
6	209.5	7.3	1203	2 C89217	protein C5SA6.2
7	166	5.8	662	2 T20343	hypothetical prote
8	161	5.6	403	2 T37571	tubulin-tyrosine 1
9	155.5	5.4	883	2 A96805	hypothetical prote
10	144.5	5.1	753	2 S48261	hypothetical prote
11	130	4.6	2472	2 E83594	still frame shift p
12	124	4.3	4660	2 T42737	gp330 protein prec
13	120	4.2	882	2 A70507	probable reductase
14	119.5	4.2	3436	2 S55659	segment protein 6
15	115.5	4.0	1006	2 T41104	probable transcrip
16	114.5	4.0	1336	2 S25716	Ras guanine nucleo
17	112.5	3.9	696	2 T46394	hypothetical prote
18	112.5	3.9	1575	2 S68448	synaptotjanin, 170K
19	111	3.9	930	2 A25923	progesterone recep
20	109	3.8	906	2 A43817	transforming prote
21	108.5	3.8	741	2 I51657	suppressor of yeas
22	108.5	3.8	1048	2 T31425	C-terminal domain-
23	108.5	3.8	1650	2 S53457	dominant autoantig
24	108.5	3.8	2176	2 T39188	probable U5 snRNP-
25	108	3.8	940	2 JE0291	FB19 protein - hum
26	108	3.8	1245	2 G84897	hypothetical prote
27	108	3.8	3149	1 QQB88	BP1F1 protein - hu
28	107.5	3.8	420	2 T46910	hypothetical prote
29	107.5	3.8	896	2 B43817	transforming prote

30	107	3.7	755	2 S32103	filensin - bovine
31	107	3.7	1234	2 T00363	hypothetical prote
32	106.5	3.7	538	2 S57459	hook-containing pr
33	106.5	3.7	2870	2 H96974	cyclic beta 1-2 g1
34	105.5	3.7	517	2 E89530	protein H28G03.2
35	105.5	3.7	4957	2 T03455	ALR protein - huma
36	105.5	3.7	5262	2 T03454	ALR protein - huma
37	104	3.6	836	2 T32298	hypothetical prote
38	103.5	3.6	1290	2 T00018	period protein hom
39	103.5	3.6	2774	2 A43359	microtubule-associ
40	103	3.6	670	2 D86176	hypothetical prote
41	103	3.6	816	2 T19049	hypothetical prote
42	102.5	3.6	775	2 S63626	homeotic protein A
43	102.5	3.6	906	2 JC5963	stable tubule only
44	102	3.6	454	2 A84162	hypothetical prote
45	102	3.6	454	2 T08280	hypothetical prote

ALIGNMENTS

RESULT 1

T12515
hypothetical protein DKFZp434B103.1 - human
C;Species: Homo sapiens (man)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C;Accession: T12515
R;Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z17524
A;Accession: T12515
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-352 <WAM>
A;Cross-references: UNIPROT:Q9Y4R7; EMBL:AL096725
A;Experimental source: adult testis; clone DKFZp434B103
C;Genetics:
A;Note: DKFZp434B103.1

Query Match 33.5%; Score 956.5; DB 2; Length 352;
Best Local Similarity 54.2%; Pred. No. 4.2e-64;
Matches 179; Conservative 55; Mismatches 91; Indels 5; Gaps 2;

QY	74	DIDTSADAVEDITEAEWEDLTQOYYSLVHGDAFISNRYFSOCQALLANRITVSNPQTDI	133
Db	2	DIDKDLAEPLYLTPEGWSLFLQRYQVHVGEALRLDTQVQRCEDILQQLQAVVPQIDM	61
QY	134	DGLRNIIWKPAKSRGRDIVCMRDVEEILEAAADHPLSRDNKRVVVKYIETPLLICDT	193
Db	62	EGDRNIWIVKPKSKRGKIMCMHLEMLKLVNGPNVVKDGKVVQKYEPLLIFFT	121
QY	194	KFDIROWFLVTDMNPLTIWFKESYLRFTQRPFLDKLSAHLCCNNAVKYLNVDVGRS	253
Db	122	KFDLRQWFLVTDMNPLTVWFYRDSYIRFTQPPSLKNLDSVHLCCNNSIKHLENSCHR	181
QY	254	PLIPAHNMVTSRFQYLRQGRGAVGWSVIVPSMKKATAHAKMVAQDHVPEKNSFELY	313
Db	182	PLLPDPNMSSORFQAHLEMGAPNAWSTIIVFGMKDAVIHALQTSQDTVQCRASFELY	241
QY	314	GADFVLGRDFRPLWIEINSPTHPTPTVTAQLCAQVQEDTIKIVAV----	369
Db	242	GADFVGEDEFQPLWIEINASPWPATATATARLCAGVQADTLRVVIDRMLDRNCDTGAFE	301
QY	370	LLWRQPVVEPPPPFSGSDCLCAGV-SVRRAR	398
Db	302	LIYQPVTTTSPASTPRPSCLLPWYSOTRAR	331

RESULT 2

A45443
tubulin-tyrosine ligase (EC 6.3.2.25) [validated] - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A45443
R;rsfeld, K.; Wehland, J.; Plessmann, U.; Dödemont, H.; Gerke, V.; Weber, K.
J. Cell Biol. 120, 725-732, 1993
A;Title: Characterization of the tubulin-tyrosine ligase.
A;Reference number: A45443; MUID:93147125; PMID:8093886
C;Accession: A45443
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-379 <ERS>
A;Cross-references: UNIPROT:P38160; GB:X68453; GB:S54050; NID:92199; PIDN:CAA48494.1; PID:3881506; GSPDB:GN00021; CESP:ZK1
A;Experimental source: brain
A;Note: Sequence extracted from NCBI backbone (NCBIN:123858, NCBIP:123859)
C;Keywords: ligase

Query Match 9.1%; Score 259; DB 2; Length 379;
Best Local Similarity 26.8%; Pred. No. 1e-11;
Matches 98; Conservative 57; Mismatches 128; Indels 82; Gaps 20;

QY 52 PG--QLVDI---ACKVC-QAYLQLEHEDIDTSADAVEDLTEAWEDLTQYYSL-----100
DB 56 PGLMQLVNYRGADKLCRKASLVLK-----IKTSPELAEST---WFPESVVIYPTNLKTP 108
QY 101 -----VHGDAFINSRNYFSQCOALLNRITSVNPQTDIDGLRNIIWIKPAKSR 149
DB 109 VAPAQNGIHPPIHSSR--TDREFF-----LTSYNKKKE-DGEGNVWIAKSAK 156
QY 150 GRDIVCMRVETLELAADHPLSRDNK---WVQKYIETPLLI--CDTKFDIROWFLVT 204
DB 157 GEGILSSSEATELDFI-----DNQGVHVQIKTERPLLECHRKFDIRSWLVD 208
QY 205 DWNPLTIWFKESYLFSTQRFSLDKL-DSAIHLCNNAVQKYLKNDVGRSPLPAHNMWT 263
DB 209 --HOYNIYLYREGVLTASEPHYTDNFQDKTCHLTWHCIQKEYSKNYGK--VEEGNEMP 263
QY 264 STRFQYLRQGRGAVGWSVIYPSMKKAIAMKVAQDHVEPR-----KNSFELYGAD 316
DB 264 FEEFNQYL-----SALNTLESSILLOIKHIRSCLLSVEPAISTRHLFPYQSFQFGFD 318
QY 317 FVLGRDFRPLWIRINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDRSCDIGNFELLWROPV 376
DB 319 FWDEDLKWLIEVNGAPAC--AQKYAELC-----QGIVDIAIASVFPDPDAE----QQQ 368
QY 377 VEPPT 381
DB 369 QQPPP 373

RESULT 3
E88575
protein ZK1128.6 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: E88575
R;Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88575
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-640 <STO>
A;Cross-references: UNIPROT:Q09647; GB:chr_III; PIDN:CAA87425.1; PID:g3881506; GSPDB:GN00021; CESP:ZK1
A;Gene: ZK1128.6
A;Map position: 3

Query Match 8.8%; Score 252.5; DB 2; Length 640;
Best Local Similarity 25.6%; Pred. No. 6.8e-11;
Matches 81; Conservative 56; Mismatches 118; Indels 61; Gaps 12;

QY 132 DIDGLRNIIWIKPAKSRGRDIVCMRVETLELAADHPLSRDNKVVQKYIETPLLIC 191

DB 283 ETDAASHV-IVKPPASARGTGISVTRKPKDFPTATL-----VAQHVIERPLTN 331
QY 192 DTKFDIROWFLVDNPLTIWFKESYLFSTQRFSLDK---LDSAIHLCNNAVQKYLK 248
DB 332 RAKFDRLRYAYPTPEPLRVYIYDQGLVRFASVPYSHSVSTISNKMHLTNTYSINKLAE 391
QY 249 D-VGRSPL--LPAHNMWTSTF-----QEYLQROGRGAVGWSVIYPSMKKAIAMKVA 299
DB 392 DGVAKPVPKWTLLHLW--EHDEMGVDREKIQRE-----IEEVIKAFISTEKPI 440
QY 300 QDH-----VEPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTI 355
DB 441 REHMSRFLQEFCYELFGIDIIIDEDYKPLWLEVNISPLSHSGTPLDVSVKAPLAKDVL 500
QY 356 KVAVDRSCDIGNFELLWROPVVEPPPF---SSDCLCVAGSVRRARRQVLP-----VCNLK 408
DB 501 NLA-----GVYVPPSFDKLSADYSTPRNGRKNGRQLEIKESVAAAYK 544
QY 409 ASASLLDAQPLKARGP 424
DB 545 DQLGVIDNRIFKRLTP 560

RESULT 4
T27699
hypothetical protein ZK1128.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27699
R;Berks, M.
submitted to the EMBL Data Library, January 1995
A;Reference number: Z20407
A;Accession: T27699
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-680 <MIL>
A;Cross-references: UNIPROT:Q09647; EMBL:247357; PIDN:CAA87425.2; GSPDB:GN00021; CESP:ZK1
A;Experimental source: clone ZK1128
C;Genetics:
A;Gene: CESP:ZK1128.6
A;Map position: 3
A;Introns: 114/1; 208/3; 242/3; 334/3; 362/2; 405/2; 424/2; 550/2; 633/2

Query Match 8.8%; Score 252.5; DB 2; Length 680;
Best Local Similarity 25.6%; Pred. No. 7.4e-11;
Matches 81; Conservative 56; Mismatches 118; Indels 61; Gaps 12;

QY 132 DIDGLRNIIWIKPAKSRGRDIVCMRVETLELAADHPLSRDNKVVQKYIETPLLIC 191
DB 323 ETDAASHV-IVKPPASARGTGISVTRKPKDFPTATL-----VAQHVIERPLTN 371
QY 192 DTKFDIROWFLVDNPLTIWFKESYLFSTQRFSLDK---LDSAIHLCNNAVQKYLK 248
DB 372 RAKFDRLRYAYPTPEPLRVYIYDQGLVRFASVPYSHSVSTISNKMHLTNTYSINKLAE 431
QY 249 D-VGRSPL--LPAHNMWTSTF-----QEYLQROGRGAVGWSVIYPSMKKAIAMKVA 299
DB 432 DGVAKPVPKWTLLHLW--EHDEMGVDREKIQRE-----IEEVIKAFISTEKPI 480
QY 300 QDH-----VEPRKNSFELYGADFVLGRDFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTI 355
DB 481 REHMSRFLQEFCYELFGIDIIIDEDYKPLWLEVNISPLSHSGTPLDVSVKAPLAKDVL 540
QY 356 KVAVDRSCDIGNFELLWROPVVEPPPF---SSDCLCVAGSVRRARRQVLP-----VCNLK 408
DB 541 NLA-----GVYVPPSFDKLSADYSTPRNGRKNGRQLEIKESVAAAYK 584
QY 409 ASASLLDAQPLKARGP 424
DB 585 DQLGVIDNRIFKRLTP 600


```
RESULT 5
T20262
hypothetical protein C55A6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20262
R:Kershaw, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19243
A:Accession: T20262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1198 <WIL>
A:Cross-references: UNIPROT:O17720; EMBL:Z81051; PIDN:CAB02862.2; GSPDB:GN00023; CESP:C55A6
A:Experimental source: clone C55A6
C:Genetics:
A:Gene: CESP:C55A6.2
A:Map position: 5
A:Introns: 32/3; 68/1; 107/3; 201/3; 273/1; 364/3; 407/3; 460/3; 476/2; 534/2; 677/3; 73
Query Match 7.3%; Score 209.5; DB 2; Length 1198;
Best Local Similarity 25.5%; Pred. No. 2.9e-07;
Matches 60; Conservative 57; Mismatches 87; Indels 31; Gaps 9;
Qy 140 WIIKPAKSGRDIVCMRVVEILEAAADHPLSRDNKWWVQKVIETPLICTKPIRQ 199
Db 715 FIVKPTNSRQKGIFFANSWADI-----PARGPL-----LVSRYLKDPYLVNNHKFDLRI 764
Qy 200 WFLVTDNPLTIWFKESYLRFSTQRE--SLDKLDS--AIHLGNNAVOK----YLKNDVG 251
Db 765 YVAVTSFPLVAVYSGLARLASRPYDTSASSADSNEYVHLTNYSINKNSTSFVRNESH 824
Qy 252 RSPLLPAHNMWTTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAMKVAQDHV--EPRKN- 308
Db 825 SSEDLA--GHKWTLLGALLRYVENEKDA---KLLMLRIEDLIVKSLLSIGNSVATASRTNL 879
Qy 309 -----SPELGCADFLVGRDRFPLWIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVA 358
Db 980 RFACITNFEFGFVLVDQALKPWLLEVNLSFSLACDAPLDSLLKTRLIADLLNLA 934
RESULT 6
C89217
protein C55A6.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: C89217
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C.elegans/
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C.elegans/
A:Accession: C89217
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1203 <STO>
A:Cross-references: UNIPROT:O17720; GB:chr_V; PIDN:CAB02862.1; PID:g3875273; GSPDB:GN00023
C:Genetics:
A:Gene: C55A6.2
A:Map position: 5
Query Match 7.3%; Score 209.5; DB 2; Length 1203;
Best Local Similarity 25.5%; Pred. No. 2.9e-07;
Matches 60; Conservative 57; Mismatches 87; Indels 31; Gaps 9;
Qy 140 WIIKPAKSGRDIVCMRVVEILEAAADHPLSRDNKWWVQKVIETPLICTKPIRQ 199
Db 720 FIVKPTNSRQKGIFFANSWADI-----PARGPL-----LVSRYLKDPYLVNNHKFDLRI 769
Qy 200 WFLVTDNPLTIWFKESYLRFSTQRE--SLDKLDS--AIHLGNNAVOK----YLKNDVG 251
Db 770 YVAVTSFPLVAVYSGLARLASRPYDTSASSADSNEYVHLTNYSINKNSTSFVRNESH 829
```

```
Qy 252 RSPLLPAHNMWTTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAMKVAQDHV--EPRKN- 308
Db 830 SSEDLA--GHKWTLLGALLRYVENEKDA---KLLMLRIEDLIVKSLLSIGNSVATASRTNL 884
Qy 309 -----SPELGCADFLVGRDRFPLWIEINSSPTMHPSTPVTQAQLCAQVQEDTIKVA 358
Db 885 RFACITNFEFGFVLVDQALKPWLLEVNLSFSLACDAPLDSLLKTRLIADLLNLA 939
```

```
RESULT 7
T20343
hypothetical protein D2013.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20343; T22085
R:Mortimore, B.
submitted to the EMBL Data Library, January 1995
A:Reference number: Z19259
A:Accession: T20343
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-662 <WIL>
A:Cross-references: UNIPROT:Q09512; EMBL:Z47808; PIDN:CAA87778.1; GSPDB:GN00020; CESP:D2013.9
A:Experimental source: clone D2013
R:Matthews, P.
submitted to the EMBL Data Library, January 1995
A:Reference number: Z19510
A:Accession: T22085
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-662 <WIL>
A:Cross-references: EMBL:Z47809; PIDN:CAA87783.1; GSPDB:GN00020; CESP:D2013.9
A:Experimental source: clone F42A8
C:Genetics:
A:Gene: CESP:D2013.9
A:Map position: 2
A:Introns: 78/2; 114/3; 330/3; 471/1; 521/2; 587/2
Query Match 5.8%; Score 166; DB 2; Length 662;
Best Local Similarity 24.3%; Pred. No. 0.00022;
Matches 72; Conservative 45; Mismatches 115; Indels 64; Gaps 14;
```

```
Qy 78 SADAVEDLTAEWEDLTQQYISL-VHGDAFISNRYFSOCALLNRITSVNPOTDIDGL 136
Db 396 AACAMRDPKNDYQLT---YNLNTQLPEFARFQN-----RELNGQ 434
```

```
Qy 137 RNIWIKPAKSGRDIVCMRVVEILEAAADHPLSRDNKWWVQKVIETPLLI----- 190
Db 435 HNWIVKPMNLARGMDMTVTEDLNQIRMI-----ETGPKIVCEYIPRPLPPRPDNG 487
```

```
Qy 191 CDTKFDIRQWFLVDNPLTIWFKESYLRFSTQRFSLDKL-DSAIHLGNNAVQKYLKND 249
Db 488 NKVKFDLRIYVFLNGAPVTAYVNFIRFALNEFSLNPFEDVETHF---TVENYL--D 542
```

```
Qy 250 VGRSPLLPAAHNMWTTSTRFOEYLQROGRGAVGWSV---IYPSMKKAI-AMKVAQDHVEP 305
Db 543 KEKILQKNCEN-----FETIEKAPRIQWSEVQKDINLTIRKAIKAEAAKEAPRGVAP 596
```

```
Qy 306 RKNSELYGADFLVGRD-----FRPLWIEINSSPTMHPSTPVTQAQLCAQVQEDTI 355
Db 597 NVQSRAMYGVIMLQHGDNNDVIKSTLLEINFMFD-----TTRACQYYPDFADTV 645
```

```
RESULT 8
T37571
tubulin-tyrosine ligase related protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T37571
R:Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z21726
```


A:Accession: T37571
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-403 <BAD>
A:Cross-references: UNIPROT:Q10438; EMBL:Z70721; PIDN:CAA94694.1; GSPDB:GN00066; SPDB:SE
A:Experimental source: strain 972h-; cosmid c12B10
C:Genetics:
A:Gene: SPDB:SPAC12B10.04
A:Map position: 1
A:Introns: 320/3; 348/3

Query Match 5.6%; Score 161; DB 2; Length 403;
Best Local Similarity 20.6%; Pred. No. 0.00025;
Matches 72; Conservative 65; Mismatches 127; Indels 86; Gaps 14;

QY 71 EHEIDTSDAV-----EDTAEWEDLTQQVYSLVHGDAFTNS-----110
DB 49 QYEDID--FDEVYKPKTKLCCSVIRKALIRKEYLWRVITYLAKHPSILSKSVPEAY 106

QY 111 ----RNYFSQCQALLNRITSVPQTDIDGLRNI-----WIIPKAASGRGRDIVCMRVERI 162
DB 107 SLELDVYAEFLDLSMEAYELRQLEENATKNISEKQWYILKPSMCDRAQGIRLFSFIEEL 166

QY 163 LEL-----AAADHPLSRDNK-----WVQKIETPLLICDTKFDI 197
DB 167 QAIFDSFDDSESESEAGLEEKGDITVAFNNKIVISQIRNFLVQKYISKPLLLIDHRKPHI 226

QY 198 RQWFLVTDWNPITWFKESYLFSTQFSLDKLDSAI---HLNNAVQYKLNKDVGRSP 254
DB 227 RAYVLAT--GALSYVLFNEMCLLARDKYKTPDPDLFLSHLSNTCLQ---GDNVEQSS 281

QY 255 LPLAHNMWTSRFOEYLQQRGAVGVSIVTSPMKKAIAHAMKVAQD-----HVEPRKNSF 310
DB 282 I---RDFWNTS-----IENK-----DDIFKSLNIIGDVFEAAATTQGIHQPLENCF 326

QY 311 ELYGADFLVGRFRPRLWIEINSPTWHPSTPTVAQLCAQVQEDTIKAVD 360
DB 327 EIFGVDFLDCESQVYLLEVS-----YPDFKQTGKNLSNIENLFSAVVE 372

RESULT 9
A96805
Hypothetical protein TSM16.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96805
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, W.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Nelson, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96805
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-883 <STO>
A:Cross-references: UNIPROT:Q9CAQ1; GB:AE0051173; NID:g3382502; PIDN:AAF07788.1; GSPDB:GN
C:Genetics:
A:Gene: TSM16.14
A:Map position: 1

Query Match 5.4%; Score 155.5; DB 2; Length 883;
Best Local Similarity 25.3%; Pred. No. 0.0021;
Matches 55; Conservative 36; Mismatches 93; Indels 33; Gaps 7;

QY 134 DGLRNWIIIPKAASGRDIVCMRVERIELEAAADHPLSRDNKVVVQKIETPLLICDT 193
DB 641 DQLNNLWILKPNMARTIDTSDNLSAIR-----MMETGPKTCQKIYIEHPALFKGN 693

QY 194 KFDIROWFLVTDWNPITWFKESYLFSTQFSLDKLDSAIHLNNAVQYKLNKDVGRS 253
DB 694 KFDLRVVLVRSIDPLEIYLIEIFWVRLSNPNVPSLEKHSFFETHTFTVMNY-----GRK 748

QY 254 PLLPAHNMWTSRFOYLRQSG-----RGAVWGSV-----IYFSMK---KAIAHAMK 297
DB 749 ----LNHKPTASFRFEQEHNDYAFHFVNNTIQLSIVKWMMDIHEKVKQVIRAVFEAAA 804

QY 298 VAQDHVEPRKNSFELYGADFLGRDPRPWLIEINSSP 334
DB 805 LAHPMQSPK--SRAMYGVDVMDLSDSPEPKILEVTYCP 840

RESULT 10
S48261
Hypothetical protein YBR094w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR0821
C:Species: Saccharomyces cerevisiae
C:Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 09-Jul-2004
C:Accession: S48261; S45962; S41800; S44676
R:Mannhaupt, G.; Stucka, R.; Ehmlé, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A:Reference number: S48255; MUID:95208357; PMID:7900426
A:Accession: S48261
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-753 <MAN>
A:Cross-references: UNIPROT:P38254; EMBL:X78993; NID:g476045; PIDN:CAA55599.1; PID:g4760;
R:Feldmann, H.; Mannhaupt, G.; Schwarziöse, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S45962
A:Molecule type: DNA
A:Residues: 1-753 <FE2>
A:Cross-references: EMBL:Z35963; NID:g536366; PIDN:CAA85047.1; PID:g536367; MIPS:YBR094w
R:Dekker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivell, L.A.
submitted to the EMBL Data Library, December 1992
A:Reference number: S31313
A:Accession: S41800
A:Molecule type: DNA
A:Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
A:Cross-references: EMBL:X69881; NID:g4086; PIDN:CAA49508.1; PID:g4088
C:Genetics:
A:Cross-references: SGB:S0000298
A:Map position: 2R

Query Match 5.1%; Score 144.5; DB 2; Length 753;
Best Local Similarity 21.8%; Pred. No. 0.011;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;

QY 140 WIIPKAASGRDIVCMRVERI-----LELAADHP-----171
DB 476 WTVKPSMSDKGQIRVFKTIEDQLAIFDSFDEDSAEESGNDDDADDVNGEFMDNNKN 535

QY 172 LGRDNKVVQKIETPLIIC---DTKFDIROWFLVTDWNPITWFKESYLFSTQF--226
DB 536 ISQLRHFIIQEVLTNPILLASMDNKKFIRCY--VVCGRDLQVVFVYDRMLALFAAKFPVP 593

QY 227 -----SLDKLDSAIHLNNAVQYKLNKDVGRSPLLPAHNMWTSRFO--EYLQQRG 277
DB 594 LDPYAYSVTDLKDLCHLTNTCLQS--KKDKDSSVL-----EFDISIEIPNERS 642

QY 278 AVWGSVYIPSMKKATAHAMKVAQDHVEPRKNSFELYGADFLGRDPRPWLIEINSSPTMH 337
DB 643 NT-KEQIHSITNDVFLAAVNVNRLNFQPLNPAFETYGVDFLIDSNYEYVKKLEINAFPDFK 701

QY 338 PSTPVTVAQLCAQVQEDTIKAV 359
DB 702 QTGKDLKNLIDELFDDTVKVCV 723

RESULT 11

E83594
still frameshift probable component of chemotactic signal transduction system PA0413 [im
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: E83594
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83594
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2472 <STO>
A;Cross-references: UNIPROT:Q91696; GB:AE004478; GB:AE004091; NID:99946261; PIDN:AAG0380
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA0413

Query Match 4.6%; Score 130; DB 2; Length 2472;
Best Local Similarity 19.7%; Pred. No. 0.76;
Matches 122; Conservative 81; Mismatches 203; Indels 214; Gaps 29;
Qy 71 EHEDITD-----SADAVEDLAEWEDLTQOYSLVHGDAFISNSRNYFSCQALLN 122
Db DHEVDAERARLAGPRDAMRSVVGALCEELVRIKDSL---DLFVRSDRGHPSLELDALLA 354
Qy 123 RITSVNPQTIDG-----LKNWIKPAKSRGR--DIVCMRVETEELEAA----- 167
Db 355 PLQIADTLAVLFGQPRKVLQDLQVHALAGRPPSDAILMDVAGALLYVEATLAGM 414
Qy 168 ---ADHPLSRDNKVVQKYIET-PLLICDTKFDIRQ-----WFLVTDNMPLTIWFYKE- 216
Db 415 AGPGDERNSESLPTTDDVAQIHQVLIKEARNGLEQAKDAIEFASQWNHEHLARVPEL 474
Qy 217 -SVLRSTORFSLDKLSAHLGNVAVQKYLKNDVGRSPLLPAHNMWTS-----TRF 267
Db 475 LTQVRGLAMIPLERAATLTETCNRIQEQL---LARK-----AVPDWQSLDTLADAITSV 527
Qy 268 QEYLQGRGAVGWSVITYPSMKKAIAHAMKVAODHVEPRKNSPELYGADPVLGRDFRPMW 327
Db 528 EYTLER-----SEDHASQSDLI---LDVAEDSLAN-----LGYTKLP-- 562
Qy 328 IEINS-----SPTMHPSTPVTQAQCAQ-----VOEDTIKVA 358
Db 563 ---NSSAPABPGLSGPAIESPAAPERPEAVVEVAETAEPADTAPAEAREADAPOLA 619
Qy 359 VDRSCDIG-----NPELLWRQPV---VEPPP-----F 382
Db 620 SDNNTLGEVAPDAGEPFLDLDLPLDSDSAEPFALPEVVESSGQPSQTSAPARSLDDF 679
Qy 383 SGSDLCVAGSVRRARRQVLPVCNKLKASALLDAQPLKARGPSAMPD----- 429
Db 680 SLDEILSLGD-----LPA-----DAAP--ASGPALADWSLPEQWGLGDDL 719
Qy 430 --PAQGPSS-----PALQDRLGLKEEKGKPLALLAPLRGAESG---GAAQP-----T 472
Db 720 AQAQAGETLDSLRLPALSFDAPLESLE--PLPALEFPDGSAAEQELVLDLALDPLDVA 777
Qy 473 RTVAAGKV-----ELPACPRHVDQAPNTGVVPAQPAKSWDPNQLNAH 516
Db 778 LPBAEGEVSAWEGSSLEELDLSLDLPLFVQLPEAAEAEPAAEALAEAPALSALAEVMAA 837
Qy 517 PLPLVLRGLTAAGALRPPPP 536
Db 838 PVQPIPPAQNVPVSLDPPP 857

RESULT 12

T42737

gp330 protein precursor - rat

N;Alternate names: megalin

C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42737
R;Saito, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A;Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of t
A;Reference number: A58173; MUID:95024033; PMID:7937880
A;Accession: T42737
A;Status: preliminary; translated from GB/EMBL/DBU
A;Molecule type: mRNA
A;Residues: 1-4660 <SAI>
A;Cross-references: UNIPROT:P98158; EMBL:L34049; NID:9561852; PID:9561853; PIDN:AAA51369.
A;Experimental source: strain Sprague-Dawley; kidney; kidney
C;Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-4660/Product: gp330 protein #status predicted <MAT>

Query Match 4.3%; Score 124; DB 2; Length 4660;
Best Local Similarity 19.7%; Pred. No. 5.4;
Matches 118; Conservative 69; Mismatches 191; Indels 222; Gaps 31;
Qy 66 YLGQLEH-EDIDTSADAVEDLAEWEDLTQOYYS-LVHGDAFISNSRNYFSCQALLNR 123
Db 4079 YLEEEHEHIQTIYDWD-----PEHIGLSVVYTVLAQGSQFGAIKRAYIPNFESGSN- 4130
Qy 124 ITSVPNTQDID-GLRNWIKP---AAKSRGRDIVCMRVETEELELAADHPLSRDNKVV 179
Db 4131 ---NPIREVDLGLK--YLMQPDGLAVDWVGRHIVYSDAKSQRIEVATLD---GRYRKWL 4181
Qy 180 VKQYIETLLICDTKFDIRQWFLVTDNMPLTIWFYKESYLRFSTQRFSLDKLSAHLN 239
Db 4182 ITTQLDPAALIA-----VNP-----KGLMFWTDQKQPKIESA----- 4215
Qy 240 NAVQYKLNKDVGRSPLLPAHNMWTSRFBQYLRQGRGAVGWSVIYPSMKKAIAHAMK-- 297
Db 4216 -----WMNGEHRSLVSENLCWPNGLSIDYLNDD-----RVYMSDSKEDVIEALKYD 4262
Qy 298 -----VAQDHVEP-----RNSFELYGADPVLGRDFRPMWLEI 330
Db 4263 GTDRRLIINEAMKPFSLDIFEDKLYWVAKEKGWVRQNKFGKENKEVL--VVPMLTQV 4320
Qy 331 N-----SSPTMHPSTPVTQAQ-----CAQVEDTIKVAVDNRSCDIGNFELLWRQ 374
Db 4321 RIFHLRYNQSVNSPCKVCCHLCLLRPGYSCACPGSDFTVTGTVQCDAS-EL----- 4375
Qy 375 PVVEPPP-----VCNKLKASALLDAQPLKARGPSAMPDPAQG----- 433
Db 4376 PVTMPPPCRMHGNCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 4429
Qy 404 -----VCNKLKASALLDAQPLKARGPSAMPDPAQG----- 433
Db 4430 LTFVIVIVGALVLVGLFHYRKTGSLTLP-LKPLSLSLAKPSNGNGVTFRSGADVNM 4488
Qy 434 -----PSPALQDRLGLKE---EKG-LPLALLAPLRGAESGGAQPTRTKAAGKVE 481
Db 4489 DTGVSPFGPETIIDRSMANNEHFVMEVGQPVIFENPMYAAKDN-----TSKVA 4537
Qy 482 LPACPCRHRVDSQAPNTGVVPAQPAKSWDPNQLNAHPLVLRGLKTAEGALRPPPGKGS 541
Db 4538 LAV-----QGPSTGAQVTPENV--ENQNYGRPIDP-----SEIYPEKPPASPGA 4580

RESULT 13

A70507

probable reductase (EC 1.3.99.-) iron-sulfur protein - Mycobacterium tuberculosis (strain
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: A70507
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70507
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-882 <COL>
A:Cross-references: UNIPROT:O33268; GB:297991; GB:AL123456; NID:g3261837; PIDN:CAB10793.
A:Experimental source: strain H37rv
C:Genetics:
A:Gene: Rv0338c
C:Keywords: oxidoreductase

Query Match 4.2%; Score 120; DB 2; Length 882;
Best Local Similarity 27.2%; Pred. No. 0.95;
Matches 43; Conservative 11; Mismatches 52; Indels 52; Gaps 5;

QY 416 AQLPKARGSNPD-----PAQPPSPALQDLDGLKEEGLPLALLPLRGAES 465
DB 736 AKQAQERAPKAAAPKAAAPVTPVEAPAEAPQAPA-----PAAAPAPVKGLGMA 782
QY 466 GGAAOPTRTKAAGKVELPACPCRHDVDS-----QAPNTGVPVAPQPAKSW 508
DB 783 AGAKRPGAKAAPTAPAPAPVKGGLGAGAKRPGAKTTPPPAGLAEPAAPQPEA 842
QY 509 DPNQLNAHLEP-----VLRGLKTAEGALRPP 535
DB 843 KQPEPAAPKPKQTDGDPAPAPAPVKGGLGIARGA-RPP 879

RESULT 14
S55659
tegument protein 64 - equine herpesvirus 2
C:Species: equine herpesvirus 2
C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C:Accession: S55659
R:Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A:Title: The DNA sequence of equine herpesvirus 2.
A:Reference number: S55594; MUID:95302501; PMID:7783207
A:Accession: S55659
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3436 <TEL>
A:Cross-references: UNIPROT:Q66666; GB:U20824; NID:g695172; PIDN:AAC13852.1; PID:g695237
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 4.2%; Score 119.5; DB 2; Length 3436;
Best Local Similarity 18.6%; Pred. No. 7.5;
Matches 134; Conservative 94; Mismatches 261; Indels 233; Gaps 29;

QY 9 VVSHQSCSRSSKPRDQREAEAGS-----SDLSSRQDAENAEAKLRLGLPQLVDIACKVC 63
DB 1919 VAATGSLPRLRHPAPGAEADTGAAUFAUSEAVKLEKSGRIRWRGVEGKLSDAHSLVA 1978
QY 64 QAYLGQLEHEDI-----DTSADAVDLTEAEWEDLTQOYYSLVHGDADFISNRNYF 114
DB 1979 RSAAGAQRKLFALQLGLRKDYAAQEQREIME-DWK-----AFVT----- 2018
QY 115 SOQALLNRITSVNPOTDIDGLRNWIWPKPAKSGRGRDIVCMRVEIELELAAADHPLSR 174
DB 2019 ---EAPINSMEDVND-----ILRAAPNEEAPEFAKKLERVREMEAEAKLEK 2064
QY 175 DNKVVQKIETPLLIC-----DTKFDIROWFLVLT--DNPLTIWFKESYLRFSRPSL 228
DB 2065 EAEEMLTQAVKRGLOCWGRIONAFNPAGGITGEDWAAVAABFQREGSTLSSTLPGLQ 2124
QY 229 DKLDSAIH-----LCNNAVQKYLKNDVGRSP-----LLP--AH-NMWTSTRFQYLRQ 275
DB 2125 SKLTDKVEAEVALLLNKVVMLPNPAPKPPAFDMLTPYRAHMFNFKLSFPLPKLNROA 2184
QY 276 RG-----AVWGSVLYPS-----MKATAHAKV---RQDHVEPRKNSFELYGADF 317

DB 2185 EAVEAKMSIQEIQIEGADVVEAVAGTPLEAPVARALRLRLRAARDEAAGLKQIDEGERAY 2244
QY 318 VLG-----RDFRPWLIE----- 329
DB 2245 VQGVGERAGEGPPAKPAEIPKLLTYEQTLSLANLPEDFKQNVQLQNETMLNLQREYL 2304
QY 330 -----INSPMTMHST--PVTAQICAQVQETIKVAV-----DRSCDIGNFELLWR- 373
DB 2305 GRVTENINSLESRAKTSRGEANARLAIVTEENLPQANVSISRRRLDKSDPVGFLEGIVD 2364
QY 374 QPVVEPPPPSGDLGVAGSVVERARRQVLPVC--NIKASASLLDQAQPLKARGSPAMPDPA 431
DB 2365 KQIVESDPYSATRESL--VWLHRTFKALLPLCPASLKRMRMELLGEEILREKG-----RG 2416
QY 432 QQPPSPALQ----- 447
DB 2417 RGAQPRGVRAHETDDVGVLTEAIELEARRVTGGAAVEGWVKRDAYRNMVEDLALRSE 2476
QY 448 -EKGL-PL-----ALLAPLRGAASGGAAQPTRTKAAGKVELPACPCH- 489
DB 2477 VEKRLGPLVEKSRREALDAPDLAVLOEAAALLLEAKTGGLDK--SAPETHERVLEIQMYL 2534
QY 490 -----VDSQAPNTGVPVAPQAKSWDPNQLNAHLEPVLRLGLKTAEGALRPPPGCK 539
DB 2535 RPKLDFLKHLYDSQRPVFEAAPLSRALYWSNGRESRENSGEVGEVGGERRGEGGGE 2594
QY 540 GS 541
DB 2595 GS 2596

RESULT 15
T41104
probable transcription factor subunit, TPR domain - fission yeast (Schizosaccharomyces
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41104
R:Purnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21964
A:Accession: T41104
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1006 <PUR>
A:Cross-references: UNIPROT:O74458; EMBL:AL031535; PIDN:CAA20753.1; GSPDB:GN00068; SPDB:5
A:Experimental source: strain 972h-; cosmid c16C4
C:Genetics:
A:Gene: SPDB:SPCC16C4.14c
A:Map position: 3

Query Match 4.0%; Score 115.5; DB 2; Length 1006;
Best Local Similarity 21.0%; Pred. No. 2.5;
Matches 92; Conservative 58; Mismatches 137; Indels 151; Gaps 24;

QY 28 EPAGSSDLSRRD-----AENAEAKLRLGLPG---QLVDIACKVCOA---YLGQLEHEDI 75
DB 3 QNGGNSYVDNSNMNETQNDTTDNFDAEMQDLNGYISEIVDEARNVSEVDAKFLG----- 55
QY 76 DTSADAVDLTEAEWEDLTQOYYSLVHGDADFISNRNYFSQOALLNRITSVNPOTDI-- 133
DB 56 DTSALQAQGL-----WSDESDYEG-----SDDESFSK-----TASRTEDDIAN 95
QY 134 -----DGLRNWIWPKPAKSRGR---DIVCMRVEIELELAAADHPLSRDNKW- 178
DB 96 EBEWENLXAVAGFRK---VRKGHKGGRVSRADMLPSVEVQQLSL--ANHLFAQSGNFD 150
QY 179 VVQKIETPLLICDQKFDIROWFLVTDWNPLTIW-FYKESYLRFSRPSLQK-----DS 233
DB 151 EAAQLAEETVRI-----DNNVIAAWMLGEGHQRQGRNGRNVIEKCLIAMWA 196
QY 234 A1HLCNNAVQKYLKNDVGRSPDLLPAHNMW-TSTRFQYLRQGRGAVWGSVLY----- 285
DB 197 AAHL-----KP--KHLELFTCAKLSLE-----FWDQADYCYNRAVS 233


```
Qy 286 -----PSMKKAI AHAMKVAQDHVEPRKNSFELYGADFVLGRDFRPWLEINSSPTWHPs 339
Db 234 AKPPKSELKKYIWNRSVLNKEHSLKAA---EGFKFL-----QS 272
Qy 340 TPVTAQLCAQVQEDTIKAVDRSCDIGNFELLWRQPVVPEPPFSGSDLCVAGSVRRARR 399
Db 273 SPYNASILKNLAETIYIKIHAPREI-LKQFEIAWKYFYQYPAPPIGNDIF----- 320
Qy 400 QVLPVCNLKASASLDAQ 417
Db 321 -DLPTLNLYAELLLLDHQ 337
```

Search completed: September 17, 2005, 09:11:22
Job time : 46 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2005, 08:38:12 ; Search time 43 Seconds
(without alignments)
939.189 Million cell updates/sec

Title: US-10-615-659-2

Perfect score: 2854

Sequence: 1 MASSILKVVYSHQSCRSSR.....LRGLTAEGALRPPPGKGKS 541

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	212	7.4	561	4	US-09-270-767-46703
2	154.5	5.4	260	4	US-09-270-767-62411
3	154.5	5.4	507	4	US-09-270-767-46787
4	151	5.3	259	4	US-09-248-796A-14603
5	144.5	5.1	753	4	US-09-538-092-31
6	126	4.4	305	4	US-09-270-767-46577
7	126	4.4	2482	4	US-09-252-991A-16967
8	125.5	4.4	802	4	US-09-823-240A-2
9	123	4.3	884	6	5208144-8
10	123	4.3	884	6	5208144-8
11	120	4.2	542	4	US-09-949-016-6778
12	120	4.2	554	4	US-09-949-016-11687
13	120	4.2	950	4	US-09-252-991A-25927
14	114.5	4.0	1319	2	US-08-290-731C-2
15	114.5	4.0	1336	2	US-08-230-731C-6
16	113.5	4.0	2074	4	US-09-491-356C-9
17	111	3.9	806	4	US-09-902-540-14992
18	110.5	3.9	560	4	US-09-252-991A-25999
19	107.5	3.8	535	4	US-09-252-991A-17140
20	107	3.7	1183	3	US-09-134-001C-3530
21	106.5	3.7	328	4	US-09-252-991A-17729
22	106.5	3.7	574	4	US-09-270-767-45812
23	106	3.7	679	4	US-09-252-991A-18857
24	106	3.7	780	4	US-09-252-991A-32892
25	105.5	3.7	542	4	US-09-252-991A-32139
26	105.5	3.7	1706	4	US-09-252-991A-31760
27	105	3.7	416	4	US-09-252-991A-32660

28	105	3.7	561	4	US-09-252-991A-23080	Sequence 23080, A
29	105	3.7	781	4	US-09-949-016-9170	Sequence 9170, AP
30	104.5	3.7	582	3	US-08-906-865-3	Sequence 3, Appli
31	104.5	3.7	582	4	US-09-129-668-3	Sequence 3, Appli
32	104	3.6	406	4	US-09-252-991A-18696	Sequence 18696, A
33	104	3.6	955	4	US-09-949-016-8369	Sequence 8369, AP
34	103.5	3.6	248	4	US-09-252-991A-21872	Sequence 21872, A
35	103.5	3.6	798	3	US-09-150-460B-8	Sequence 8, Appli
36	103.5	3.6	875	3	US-09-150-460B-7	Sequence 7, Appli
37	103.5	3.6	1290	3	US-09-150-460B-6	Sequence 6, Appli
38	103	3.6	420	4	US-09-252-991A-20033	Sequence 20033, A
39	103	3.6	515	4	US-09-252-991A-32630	Sequence 32630, A
40	102.5	3.6	398	4	US-09-902-540-12305	Sequence 12305, A
41	102.5	3.6	1076	4	US-09-949-016-7421	Sequence 7421, AP
42	102	3.6	55	4	US-09-270-767-62317	Sequence 62317, A
43	102	3.6	809	4	US-09-252-991A-31759	Sequence 31759, A
44	102	3.6	3122	4	US-10-237-551-201	Sequence 201, App
45	102	3.6	3122	4	US-10-237-551-250	Sequence 250, App

ALIGNMENTS

RESULT 1

US-09-270-767-46703
; Sequence 46703, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46703
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46703

Query Match 7.4%; Score 212; DB 4; Length 561;
Best Local Similarity 29.2%; Pred No. 1.5e-11;
Matches 57; Conservative 37; Mismatches 75; Indels 26; Gaps 6;

Qy	179	VVKYIETPLLCIDTKFDIRQWFLVTDWNP	LTTFWYKESYLFRTQRF--LDKL-DSAI	235
Db	1	VVSKIYVPLCIDGKCDLRVYVLTSTF	DLIIYVEEGIVRLATVKYDRHADNLWPC	60
Qy	236	HLCNNAVQKYLKN-----DVGRSPL	LPAHNMWSTTRFOEYLORQG---RGAVWG-	281
Db	61	HLCNYSINKYSHSYIRSSDAQDE	DVG-----HKWTL\$ALLRHLKQSCDTRQLM	112
Qy	282	-SVIYSPMKKIAAHAMKVAQDHVEPR	KNSPELYGADPVLGRDPRPWLIEINSSPT	340
Db	113	EDLIIRKAVLACASIIISACRMFV	PNCGNCPGLYFDLIDNAHKPMLLEINLSP	172
Qy	341	PVTAQLCAQVQEDTI	355	
Db	173	PLTKVKSCLMADLL	187	

RESULT 2

US-09-270-767-62411
; Sequence 62411, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517


```

; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14603
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14603

Query Match          5.3%; Score 151; DB 4; Length 259;
Best Local Similarity 22.6%; Pred.No. 3.7e-06;
Matches 53; Conservative 50; Mismatches 81; Indels 50; Gaps 9;

Qy 132 DIQGLRNWIIPAAKSRGRDI---VCMORVEBIL-----ELAAADHP-----L 172
Db 20 EINKEEKLWIKPMSDKGGIRFTKLDQLQIFNSFEENDENEDEEGVDDEENGILL 79
Qy 173 SRDNKVVQVQIETPLLICDTKEDIQWEL---VTDWNPLOTIWFYKESYLRSTORFSL- 228
Db 80 SOLRHFIQVQYKSNPLLL--SKYDHKKFHLRTVVCGDLKVFPYKNVLTLPAGEPYKLP 137
Qy 229 ----DKLDSAIHLCCNNAVKYLNQVGRSPILLPAHNMWTTSTRFEYLQRQGRGAVMGSVI 284
Db 138 GDEDEVVSLAGHLTNTCLQE-----NEOPLV-----VPFWKLQGLADNDKNIV 180
Qy 285 YPSM----KKAIAHAKMVAQDHVPRKNSFELGYADFVLGRDRPRPWLIEINSSP 334
Db 181 FEQICDITKELFKAATSVDRKQNFQINNAIEFGVDPLVNSDFSVNLLEVNSYP 234

RESULT 5
US-09-538-092-31
; Sequence 31, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqFormatter Version 0.9
; SEQ ID NO 31
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YBR094W
US-09-538-092-31

Query Match          5.1%; Score 144.5; DB 4; Length 753;
Best Local Similarity 21.8%; Pred.No. 8.8e-05;
Matches 57; Conservative 44; Mismatches 105; Indels 56; Gaps 9;

Qy 140 WIIPAAKSRGRDIVCMORVEEI-----LELAADHP-----L 171
Db 476 WIVKPSMSDKGGIRVFKEFTIEDLQAFDSFDDSEAEESGNDDDADDVNGEFMDNKNVN 535
Qy 172 LSRDNKVVQVQIETPLLIC---DTKFDIRQWFLVTDWNPLOTIWFYKESYLRSTORF-- 226
Db 536 ISQLRHFIIQEIYTNPLLLASMDNRRFHRCY--VVCRGDLQVFVYDRMLALFAAKFPVP 593
Qy 227 -----SLDKLDSATHLCNNNAVKYLNQVGRSPILLPAHNMWTTSTRFQ--EYLQRQGRG 277
Db 594 LDPYAVSVTDKLECHLTNTCLQS--KKDKOSSVL-----EFGSIEIIPNKRKS 642
Qy 278 AVMGSVIYPSMKAIAAHAKMVAQDHVPRKNSFELGYADFVLGRDRPRPWLIEINSSPTMH 337

```


Db 643 NI-KEQHSITNDVFLAANNVRLNLFQPLNFAFTYGVDFLSDNSVEVKLEINAFDPFK 701
Qy 338 PSTVTAQLCAQVQEDTIKAV 359
Db 702 QTGDKLNLIDELFDDTKYCV 723

RESULT 6

US-09-270-767-46577
; Sequence 46577, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46577
; LENGTH: 305
; TYPE: PRT
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-46577

Query Match 4.4%; Score 126; DB 4; Length 305;
Best Local Similarity 23.6%; Pred. No. 0.0013;
Matches 42; Conservative 36; Mismatches 62; Indels 38; Gaps 8;

Qy 180 VOKIETPLICDPKDIQWFLVDNPLTINFWKESYL-RFTQRFSLDKLDSAIHL 238
Db 1 VQEVQRPYLVDGHKFDIGVTVITSNPLRVIIYTGDLFRYCPVKYHFDKEN----- 55
Qy 239 NNAVQKYLKNDVGRSPLPA-----HNMW---TSTRFOEYLQRCR--CAVWGSKI 284
Db 56 ---VDKIIVGD---DVLPTWEVPSLRKYNRFGSGNRTVFEAYVRDQKDPQIWPQVE 108
Qy 285 Y-----PSMKKAIAMKVAQDHVEPRKNSFELYGADFVLGRDRFWLWLEINSSPTM 336
Db 109 HIVRTTIAAKEKDIVNLRSYRTH-----NFFDLMRFDLFDLKVFLMEANMSPNL 161

RESULT 7

US-09-252-991A-16967
; Sequence 16967, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16967
; LENGTH: 2482
; TYPE: PRT
; ORGANISM: *Pseudomonas aeruginosa*
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1766)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-16967

Query Match 4.4%; Score 126; DB 4; Length 2482;
Best Local Similarity 19.7%; Pred. No. 0.038;
Matches 122; Conservative 79; Mismatches 207; Indels 210; Gaps 29;

Qy 71 EHEDIDT-----SADAVEDLTEAWEDLTQOYYSVLVHGDAFISNSRNYFSOCOALLN 122
Db 308 DHTVDAERARLAGPRDAMRSVVGALCEBELVRIKDSL-----DLFVRSDRGHPELDALLA 364
Qy 123 RITSVNPQTDIDG-----LRNIWIIKPAKSRGR--DIVCMKRVVEILELAA----- 167
Db 365 PLQIADTLAVLFGQPRKVILOQDVIHALAQCRREPSDAILMDVAGALLYVEATLAGM 424
Qy 168 ---ADHPLSRDNKVVQKIET-PLLICDTKFDIQQ-----WFLVTDWNPPLTIWFYKE- 216
Db 425 AGPGDERNSEESRLPTTVDVAIQHQLVKEARNGLEQAKDAIIEFIASQMNHEHLARVPEL 484
Qy 217 -SYLRFSTQRFSLDKLDSAIHLNNAVQKYLKNDVGRSPLLPANMMWTS-----TRF 267
Db 485 LTVRGGLAMIPERATLTLETNRYIQEQL-----LARK-----AVPDHQSILDTLADAITSV 537
Qy 268 QEYLQRCRGAVWGSVIYPSMKKAIAMKVAQDHVEPRKNSFELYGADFVLGRDRFWL 327
Db 538 EYLERL-----SEDHASQSDLI--LDVAEDSLAN-----LGYTLKP-- 572
Qy 328 IEINS-----SPTMHPSTPVTALCAQ-----VQEDTIKVA 358
Db 573 ---NSSAPABPGLSGPAIESPAAPERPEAVVEVAETAEOPPADTAPAEAREDAFOLA 629
Qy 359 VDRSCDIG-----NPELLWRQPV---VEPPP-----F 382
Db 630 SDNWTLGEVAPDAGPFSLDLALDLDDSAEVPALPEVVEESGQPSQSTPAPARSLDDF 689
Qy 383 SGSDLCVAGSVRRARRQVLVPCNLKASALLDAQPLKARGPSAMPD---PAQ----- 432
Db 690 SLDEIDLGLD-----LPA-----DAAP--ASGPAALADWSLPEQWGLGDDL 729
Qy 433 -GPSPALQDLDLKE-----EKGLPLALLAPLRGAESG-----GAAQP-----TRT 474
Db 730 AQPTQAGETIDLSLEBPALSFDAPLESLEPLPALEPFDGSAEQELVLDALDPLDLVALP 789
Qy 475 KAAGKV-----ELPACPCRHVDQAPNTGVPVQAQPAKSWDPNQLNAHPL 518
Db 790 EAGEVSAWEGSLEELDLSLDLPEVQLPEAEAEAPPAEALASEAPALSLEAVMAAPV 849
Qy 519 EPLVRLKTAEGALRPPP 536
Db 850 QPINPPAQNVVPSLLPPP 867

RESULT 8

US-09-823-240A-2
; Sequence 2, Application US/09823240A
; Patent No. 6716597
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; TITLE OF INVENTION: Motility
; FILE REFERENCE: M00656.70064.US
; CURRENT APPLICATION NUMBER: US/09/823,240A
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: *Mus musculus*
US-09-823-240A-2

Query Match 4.4%; Score 125.5; DB 4; Length 802;
Best Local Similarity 24.7%; Pred. No. 0.0068;
Matches 78; Conservative 15; Mismatches 110; Indels 113; Gaps 15;


```
QY 264 STRFOYLOQRGAVWGSIYPSMKAIAMKVAQDHVPRKNSFELYGADFLGRDP 323
Db 322 ATRFATSL-----GSAFHP-----VLPHYATV-----PRP-----LNKNS 351
QY 324 RPLWLEINSSPTMHPST-PVTAOLCAQVQEDTTIKVAVDRSCDIGNFELLWRQP--VVEPP 380
Db 352 RP-----SSPVNTPSQPPAAKSCA-----WPTSNSPLPPSPIMISSP 391
QY 381 PFGSDLCVAGSVRRARQVLPVC-----NLKASASLLDDAOPLKARGPS 425
Db 392 PGKAT-----GPREVLFCVSSVPVQMPSPPTAENGSLDSVTYVPSPPTSGPA 440
QY 426 AMPDPAQGPSPALQRDLGLKEKGJPLALLAPLRGAESGGAQAP-----471
Db 441 APPPPPPPPPP-----PPPLPPPPPLPLASLSHCGSQASPPPGTPLASTSSKP 491
QY 472 -----TRTKAAGKVELPACPCRHVDSOAPNTGVPVA-OPAKSWDPNQL-----NAHPLEPV 521
Db 492 SVLPSPSAGAPASAEPLNP-ELGDSASSEPGLOAASQPAESPTQGLVLGPPAPPPPP 550
QY 522 LRGLKTAEGALRPPPG 537
Db 551 LPSPGAYASALPPPG 566

RESULT 9
5208144-8
; Patent No. 5208144
; APPLICANT: SMITH, JOHN A.; RAYCHOWDHURY, RAKTIMA; NILES, JOHN L.
; TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
; CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 42
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/396,697
; FILING DATE: 22-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 313,682
; FILING DATE: 22-FEB-1989
; APPLICATION NUMBER: 235,211
; FILING DATE: 23-AUG-1988
; SEQ ID NO: 8:
; LENGTH: 884
5208144-8

Query Match 4.3%; Score 123; DB 6; Length 884;
Best Local Similarity 19.7%; Pred. No. 0.014;
Matches 118; Conservative 69; Mismatches 191; Indels 222; Gaps 31;

QY 66 YLGQLEH-EDITSADAVEDLTEAEWEDLTQQYYS-LVHGDAFISNRYFSQCQALLNR 123
Db 328 YLEEEHITQIDYDWD-----PEHIGLSVYVYTVLAQGSQFGAIKRAYIPNFESGSN- 379
QY 124 ITSVPNQTID-GLRNIWIKP---AAKSRGRDIVCMRVEEILELAAADHPLSRDNKW 179
Db 380 ----NPIREVDLGLK--YLMQPDGLAVDWGRHIYWSANDANSQRIEATLD---GRYRKWL 430
QY 180 VQKYIETPLLICDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLNC 239
Db 431 ITTOLDQPAALIA-----VNP-----KLGIMFWTDGQKPKIESA-----464
QY 240 NAVQYKLVNDVGRSPLLPAAHNMWTSTRFOYLQQRGAVWGSIYPSMKAIAMK-- 297
Db 465 -----WMGEHRSVLVSENGLWPNGLSIDYLNDD-----RVYWSDSKEDVIEAIKYD 511
QY 298 -----VAQDHVEP-----RKNSFELYGADFLGRDPRPMLIEI 330
Db 512 GTDRRLIINEAMKPFSLDIFEDKLYWAKEKEGVWRQNKFKENKEKVL--VVPNMLTVQ 569
QY 331 N-----SSPTMHPSTPVTAQI-----CAQVQEDTIKAVDRSCDIGNFELLWRQ 374
Db 570 RIFHQLRYNQSIVSNPCKQVCSHLCLLRPGGYSACPCQGSDFVTGSTVQCDAS-EL-----624
QY 375 PVVEPPP-----FSGDLCVAGSVRRARRQVLP-----403
Db 625 PVTMPPPCBMGGNGCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 678
QY 404 -----VCNLKASASLLDDAOPLKARGPSAMPDPAQG-----433
Db 679 LTFVIVIIIGALVVLGFLHYRKTGSLPTLP-KLPSSLAKPSENGNGVTFRSGADVNM 737
QY 434 -----PPSPALQORDLGLKE-----EKG-LPLALLAPLRGAESGGAQOPTRTKAAGKVE 481
Db 738 DIGVSPFGPETIIDRSMAHNEHFVMEVGKQVIFENPMYAAKDN-----TSKVA 786
QY 482 LPACPCRHVDSOAPNTGVPVAOPAKSWDPNQLNAHPLEVLRGLKTAEGALRPPPGKGS 541
Db 787 LAV-----QGPSTGAQVTPENV--ENQNYGRPIDP-----SEIVPEPKPASPGA 829

RESULT 10
5208144-8
; Patent No. 5208144
; APPLICANT: SMITH, JOHN A.; RAYCHOWDHURY, RAKTIMA; NILES, JOHN L.
; TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
; CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 42
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/396,697
; FILING DATE: 22-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 313,682
; FILING DATE: 22-FEB-1989
; APPLICATION NUMBER: 235,211
; FILING DATE: 23-AUG-1988
; SEQ ID NO: 8:
; LENGTH: 884
5208144-8

Query Match 4.3%; Score 123; DB 6; Length 884;
Best Local Similarity 19.7%; Pred. No. 0.014;
Matches 118; Conservative 69; Mismatches 191; Indels 222; Gaps 31;

QY 66 YLGQLEH-EDITSADAVEDLTEAEWEDLTQQYYS-LVHGDAFISNRYFSQCQALLNR 123
Db 328 YLEEEHITQIDYDWD-----PEHIGLSVYVYTVLAQGSQFGAIKRAYIPNFESGSN- 379
QY 124 ITSVPNQTID-GLRNIWIKP---AAKSRGRDIVCMRVEEILELAAADHPLSRDNKW 179
Db 380 ----NPIREVDLGLK--YLMQPDGLAVDWGRHIYWSANDANSQRIEATLD---GRYRKWL 430
QY 180 VQKYIETPLLICDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLNC 239
Db 431 ITTOLDQPAALIA-----VNP-----KLGIMFWTDGQKPKIESA-----464
QY 240 NAVQYKLVNDVGRSPLLPAAHNMWTSTRFOYLQQRGAVWGSIYPSMKAIAMK-- 297
Db 465 -----WMGEHRSVLVSENGLWPNGLSIDYLNDD-----RVYWSDSKEDVIEAIKYD 511
QY 298 -----VAQDHVEP-----RKNSFELYGADFLGRDPRPMLIEI 330
Db 512 GTDRRLIINEAMKPFSLDIFEDKLYWAKEKEGVWRQNKFKENKEKVL--VVPNMLTVQ 569
QY 331 N-----SSPTMHPSTPVTAQI-----CAQVQEDTIKAVDRSCDIGNFELLWRQ 374
Db 570 RIFHQLRYNQSIVSNPCKQVCSHLCLLRPGGYSACPCQGSDFVTGSTVQCDAS-EL-----624
QY 375 PVVEPPP-----FSGDLCVAGSVRRARRQVLP-----403
Db 625 PVTMPPPCBMGGNGCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 678
QY 404 -----VCNLKASASLLDDAOPLKARGPSAMPDPAQG-----433
Db 679 LTFVIVIIIGALVVLGFLHYRKTGSLPTLP-KLPSSLAKPSENGNGVTFRSGADVNM 737
QY 434 -----PPSPALQORDLGLKE-----EKG-LPLALLAPLRGAESGGAQOPTRTKAAGKVE 481
Db 738 DIGVSPFGPETIIDRSMAHNEHFVMEVGKQVIFENPMYAAKDN-----TSKVA 786
QY 482 LPACPCRHVDSOAPNTGVPVAOPAKSWDPNQLNAHPLEVLRGLKTAEGALRPPPGKGS 541
Db 787 LAV-----QGPSTGAQVTPENV--ENQNYGRPIDP-----SEIVPEPKPASPGA 829
```

```
Db 625 PVTMPPPCBMGGNGCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 678
QY 404 -----VCNLKASASLLDDAOPLKARGPSAMPDPAQG-----433
Db 679 LTFVIVIIIGALVVLGFLHYRKTGSLPTLP-KLPSSLAKPSENGNGVTFRSGADVNM 737
QY 434 -----PPSPALQORDLGLKE-----EKG-LPLALLAPLRGAESGGAQOPTRTKAAGKVE 481
Db 738 DIGVSPFGPETIIDRSMAHNEHFVMEVGKQVIFENPMYAAKDN-----TSKVA 786
QY 482 LPACPCRHVDSOAPNTGVPVAOPAKSWDPNQLNAHPLEVLRGLKTAEGALRPPPGKGS 541
Db 787 LAV-----QGPSTGAQVTPENV--ENQNYGRPIDP-----SEIVPEPKPASPGA 829

RESULT 10
5208144-8
; Patent No. 5208144
; APPLICANT: SMITH, JOHN A.; RAYCHOWDHURY, RAKTIMA; NILES, JOHN L.
; TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
; CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 42
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/396,697
; FILING DATE: 22-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 313,682
; FILING DATE: 22-FEB-1989
; APPLICATION NUMBER: 235,211
; FILING DATE: 23-AUG-1988
; SEQ ID NO: 8:
; LENGTH: 884
5208144-8

Query Match 4.3%; Score 123; DB 6; Length 884;
Best Local Similarity 19.7%; Pred. No. 0.014;
Matches 118; Conservative 69; Mismatches 191; Indels 222; Gaps 31;

QY 66 YLGQLEH-EDITSADAVEDLTEAEWEDLTQQYYS-LVHGDAFISNRYFSQCQALLNR 123
Db 328 YLEEEHITQIDYDWD-----PEHIGLSVYVYTVLAQGSQFGAIKRAYIPNFESGSN- 379
QY 124 ITSVPNQTID-GLRNIWIKP---AAKSRGRDIVCMRVEEILELAAADHPLSRDNKW 179
Db 380 ----NPIREVDLGLK--YLMQPDGLAVDWGRHIYWSANDANSQRIEATLD---GRYRKWL 430
QY 180 VQKYIETPLLICDTKFDIRQWFLVTDWNPPLTIWFYKESYLRFSTQRFSLDKLDSAIHLNC 239
Db 431 ITTOLDQPAALIA-----VNP-----KLGIMFWTDGQKPKIESA-----464
QY 240 NAVQYKLVNDVGRSPLLPAAHNMWTSTRFOYLQQRGAVWGSIYPSMKAIAMK-- 297
Db 465 -----WMGEHRSVLVSENGLWPNGLSIDYLNDD-----RVYWSDSKEDVIEAIKYD 511
QY 298 -----VAQDHVEP-----RKNSFELYGADFLGRDPRPMLIEI 330
Db 512 GTDRRLIINEAMKPFSLDIFEDKLYWAKEKEGVWRQNKFKENKEKVL--VVPNMLTVQ 569
QY 331 N-----SSPTMHPSTPVTAQI-----CAQVQEDTIKAVDRSCDIGNFELLWRQ 374
Db 570 RIFHQLRYNQSIVSNPCKQVCSHLCLLRPGGYSACPCQGSDFVTGSTVQCDAS-EL-----624
QY 375 PVVEPPP-----FSGDLCVAGSVRRARRQVLP-----403
Db 625 PVTMPPPCBMGGNGCYFDENELPKCKSSGYSG-EYCEVGLS-----RGIPPGTTMAVL 678
QY 404 -----VCNLKASASLLDDAOPLKARGPSAMPDPAQG-----433
Db 679 LTFVIVIIIGALVVLGFLHYRKTGSLPTLP-KLPSSLAKPSENGNGVTFRSGADVNM 737
QY 434 -----PPSPALQORDLGLKE-----EKG-LPLALLAPLRGAESGGAQOPTRTKAAGKVE 481
Db 738 DIGVSPFGPETIIDRSMAHNEHFVMEVGKQVIFENPMYAAKDN-----TSKVA 786
QY 482 LPACPCRHVDSOAPNTGVPVAOPAKSWDPNQLNAHPLEVLRGLKTAEGALRPPPGKGS 541
Db 787 LAV-----QGPSTGAQVTPENV--ENQNYGRPIDP-----SEIVPEPKPASPGA 829
```


Db 738 DIGVSPFGPETIIDRSAMNHFVMEVGKQVPVFENPMYAAKDN-----TSKVA 786
Qy 482 LPACPCRHVDQANTGVVPAQPAKSWDPNQLNAHLEPVLRLGKLTAEAGLRPPPGKGKS 541
Db 787 LAV-----QGPSTGAQVTPENV--ENQYGRPIDP-----SEIVPEPKPASPGA 829

RESULT 11

US-09-949-016-6778
; Sequence 6778, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6778
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6778

Query Match 4.2%; Score 120; DB 4; Length 542;
Best Local Similarity 22.5%; Pred. No. 0.012;
Matches 75; Conservative 38; Mismatches 124; Indels 96; Gaps 18;

Qy 254 PLLAHNMWTSRFEYLQQRGAVGWSVIYPSMKKAIHAHMKVAQDHVEPRKNSPELY 313
Db 233 PLLPGNDV-----GRSS-YGAM--QVKQVFDYAYIVLSHAVSPLARSYPNR 275
Qy 314 GADFLVGR-----DFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCD 364
Db 276 DAESTLGRILIKVTQVEIDYRRWIKWKWSKA-HPSPGMDSRI--KIKE---RIATCNQEQ 329
Qy 365 IGNFELLWRQPVVEPP-----PFGSDLCVAGSVRRARR-----QVLPVCNLKA- 409
Db 330 TQN-----REP--ESPYGQRLTSLSSPQLSSGSSASSVSSLSGSDVDSDTPTTPTSV 382
Qy 410 -SASLLDAQPLKARGSPAMPDPAQPPSPALQRLDGLKEEKLPLALLAPLRGAA-----E 464
Db 383 YQFSLOAPAPLMAGLTALPMPS-GKPOPTTSRTLTMTNNQTRFTTIPPTLGVAPVPCR 441
Qy 465 SGGAAPQTRTKAGKVELPACPCRHVDQAPNTGVVPAQPAKSWDPNQLNAHLEPVLRLG 524
Db 442 QAGVEGTASLKAVHMSSPAIP-----SASPN---PLSSP-----HLVHKQHNG 482
Qy 525 LK-----TAEGALRPPPGKGK 540
Db 483 MKLSMKGSHGHTGGGYSSVSGGVRRPPVGNRG 515

RESULT 12

US-09-949-016-11687
; Sequence 11687, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11687
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11687

Query Match 4.2%; Score 120; DB 4; Length 554;
Best Local Similarity 22.5%; Pred. No. 0.013;
Matches 75; Conservative 38; Mismatches 124; Indels 96; Gaps 18;

Qy 254 PLLAHNMWTSRFEYLQQRGAVGWSVIYPSMKKAIHAHMKVAQDHVEPRKNSPELY 313
Db 245 PLLPGNDV-----GRSS-YGAM--QVKQVFDYAYIVLSHAVSPLARSYPNR 287
Qy 314 GADFLVGR-----DFRPWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDKSCD 364
Db 288 DAESTLGRILIKVTQVEIDYRRWIKWKWSKA-HPSPGMDSRI--KIKE---RIATCNQEQ 341
Qy 365 IGNFELLWRQPVVEPP-----PFGSDLCVAGSVRRARR-----QVLPVCNLKA- 409
Db 342 TQN-----REP--ESPYGQRLTSLSSPQLSSGSSASSVSSLSGSDVDSDTPTTPTSV 394
Qy 410 -SASLLDAQPLKARGSPAMPDPAQPPSPALQRLDGLKEEKLPLALLAPLRGAA-----E 464
Db 395 YQFSLOAPAPLMAGLTALPMPS-GKPOPTTSRTLTMTNNQTRFTTIPPTLGVAPVPCR 453
Qy 465 SGGAAPQTRTKAGKVELPACPCRHVDQAPNTGVVPAQPAKSWDPNQLNAHLEPVLRLG 524
Db 454 QAGVEGTASLKAVHMSSPAIP-----SASPN---PLSSP-----HLVHKQHNG 494
Qy 525 LK-----TAEGALRPPPGKGK 540
Db 495 MKLSMKGSHGHTGGGYSSVSGGVRRPPVGNRG 527

RESULT 13

US-09-252-991A-25927
; Sequence 25927, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25927
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25927

Query Match 4.2%; Score 120; DB 4; Length 950;
Best Local Similarity 25.1%; Pred. No. 0.031;
Matches 53; Conservative 17; Mismatches 63; Indels 78; Gaps 9;

Qy 361 RSCDIGNFELLWRQPVVEPPPF---SGSDLCVAGSVR-----RARRQVLPVCNLKA 409
Db 728 RTADVS-----RQPLRQPPALLQRRGVDLDASLRPPGFFGAARALLSPACRRGR 781
Qy 410 SASLLDAQP-----LKARGPSAM-----PDPAQG-----PPSPALQRLDGLK 446

Db 782 ATGLGQFQGRGERLSRCFGAATAAGPGCTAHARGPGGGLRRRHPPAPARPGFGLA 841
QY 447 EEKGLPLALLALRGAESGGAQPTRTKAAGKVELPACPCRHVDSOAPNTGVPVQAQPAK 506
Db 842 P-----APAGGADPGAAQPGRWPA-----PAARRQR 869
QY 507 SWDPNQLNAHPLRPVLRLGLKTABGALRPPPG 537
Db 870 TWTTRRRS-----LRTLERPAGVPAPPG 894

RESULT 14
US-08-290-731C-2
; Sequence 2, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
; TITLE OF INVENTION: AND MSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PLO921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-2

Query Match 4.0%; Score 114.5; DB 2; Length 1319;
Best Local Similarity 19.2%; Pred. No. 0.18;
Matches 118; Conservative 82; Mismatches 204; Indels 209; Gaps 31;

QY 62 VCOAYLGQLEHEDIDTSADAVEDLTAEWEDLTQYYSVLVHGDAFTNSRNPFSQOALL 121
Db 692 VCRHW---VEHFYDFERDA--DLLQR-----MEEFITGVGRAM---KKWVESITKII 737
QY 122 NR-----ITSVNPQTDIDGLRNWIITKPAKSRGRDIVCMQDRVVEETLELAAD 169
Db 738 QRKKIARDNGPGHNITFQSSPPTVE-----WHLSRFGHETFTDLLLHPIETARQLTLE 792
QY 170 HPLSR-----DNKVVVQ-KYIETPLLICDTKFDIQRQWFLVDWNPLTIWPK----- 215

Db 793 SDLYRAVQPSLVGSVWTKEDKEINSPNLL-----KMRHTTNLTLMFEKCIIVET 842
QY 216 -----ESYLRP-----STORESLDK-----LD 232
Db 843 ENLEERVAVVSRRIIEILOVFQELNNPVGVLVVSMNSPVTRLDHTTFQIISRQKKILE 902
QY 233 SAHLCNNAVOKYL-KNDVGRSPLLPAHNMW-----TSTRFQEYLQROGRGAVMGSVIY 285
Db 903 EAHLESEDHVKKYLAKLRSINPCVPFFGIGYLTNLTKEGNEPVLRRHGK-----ELIN 957
QY 286 PNMKKAIA-----HAMKVAQD-----HVEPRKNSFELYGADFVLGRDPRFW 326
Db 958 FSKRRRVAEITGEIQYQYQYCLRVPEPIKRAFFENLNPNGNSMEKEFTDYLFNKS----- 1013
QY 327 LIEIN-----SSPTMEPST--PVTAQLCAQVQEDTIKVAVDR----- 361
Db 1014 -LEIBPRHPKPLRPFPKKYSYPLKSGVRRSPRPGTMRHPTFLOQEPKIKISVSRTPES 1072
QY 362 ---SCDIGNFELLWRQPVVVEPPFFSG-----SDLCVAGSVRRARRQVLPVCLNKASASLL 414
Db 1073 TESTASAPNSP---RTPLT-PPPASGTSNTDVC---SVFSDSHSASPFHSRSASVSSI 1124
QY 415 DAQPLKARQPSAMPDPAQPPSPALQORDLGLKEEKLPLALLAPLRGAESGGAQPTRT 474
Db 1125 SL-----SKGTDEVPVP---PPVPPRRR-----PESAPAES-----SPSKI 1157
QY 475 KAAGKVELPACPCRHVDSOAPNTGVPVQAQPAKSWDPNQLNAHPLRPVLRLGLKTAE 530
Db 1158 MSKHLDSPPAIPRQPTSKAYSISDRTSISDPPE--SPPLLPPREPVRTPDVFSSS 1215
QY 531 --ALRPPPGGKGS 541
Db 1216 PLHLQPPPLGKKS 1228

RESULT 15
US-08-290-731C-6
; Sequence 6, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
; TITLE OF INVENTION: AND MSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PLO921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-2

Query Match 4.0%; Score 114.5; DB 2; Length 1319;
Best Local Similarity 19.2%; Pred. No. 0.18;
Matches 118; Conservative 82; Mismatches 204; Indels 209; Gaps 31;

QY 62 VCOAYLGQLEHEDIDTSADAVEDLTAEWEDLTQYYSVLVHGDAFTNSRNPFSQOALL 121
Db 692 VCRHW---VEHFYDFERDA--DLLQR-----MEEFITGVGRAM---KKWVESITKII 737
QY 122 NR-----ITSVNPQTDIDGLRNWIITKPAKSRGRDIVCMQDRVVEETLELAAD 169
Db 738 QRKKIARDNGPGHNITFQSSPPTVE-----WHLSRFGHETFTDLLLHPIETARQLTLE 792
QY 170 HPLSR-----DNKVVVQ-KYIETPLLICDTKFDIQRQWFLVDWNPLTIWPK----- 215

Search completed: September 17, 2005, 08:52:07
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2005, 08:48:28 ; Search time 169 Seconds

(without alignments)
1296.247 Million cell updates/sec

Title: US-10-615-659-2

Perfect score: 2854

Sequence: 1 MASSILKVVVSHQSCSRSSR.....LRGLKTAEGALRPPPGKGKS 541

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1812044 seqs, 40492589 residues

Total number of hits satisfying chosen parameters: 1812044

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2854	100.0	541	16	US-10-615-659-2
2	2854	100.0	541	16	US-10-635-977-2
3	2843	99.6	541	16	US-10-615-659-13
4	2843	99.6	541	16	US-10-635-977-13
5	1560	54.7	293	16	US-10-615-659-14
6	1560	54.7	293	16	US-10-635-977-14
7	1299	45.5	242	16	US-10-615-659-4
8	1299	45.5	242	16	US-10-635-977-4
9	1013	35.5	488	15	US-10-275-959A-5
10	969.5	34.0	362	9	US-09-925-298-617
11	969.5	34.0	362	14	US-10-102-806-617

12	956.5	33.5	352	16	US-10-615-659-7	Sequence 7, Appli
13	956.5	33.5	352	16	US-10-635-977-7	Sequence 7, Appli
14	956.5	33.5	352	18	US-10-756-143-5267	Sequence 5267, Ap
15	944.5	33.1	326	15	US-10-108-260A-4209	Sequence 4209, Ap
16	865.5	30.3	292	16	US-10-615-659-5	Sequence 5, Appli
17	865.5	30.3	292	16	US-10-635-977-5	Sequence 5, Appli
18	375.5	13.2	566	18	US-10-450-763-36330	Sequence 36330, A
19	322	11.3	330	16	US-10-615-659-26	Sequence 26, Appl
20	322	11.3	330	16	US-10-635-977-26	Sequence 26, Appl
21	303.5	10.6	1226	18	US-10-756-149-5184	Sequence 5184, Ap
22	302	10.6	268	15	US-10-424-599-205823	Sequence 205823,
23	299.5	10.5	423	16	US-10-615-659-8	Sequence 8, Appli
24	299.5	10.5	423	16	US-10-635-977-8	Sequence 8, Appli
25	280.5	9.8	439	15	US-10-104-047-3658	Sequence 3658, Ap
26	274	9.6	49	16	US-10-615-659-21	Sequence 21, Appl
27	274	9.6	49	16	US-10-635-977-21	Sequence 21, Appl
28	274	9.6	49	16	US-10-635-977-21	Sequence 22, Appl
29	274	9.6	49	16	US-10-635-977-21	Sequence 22, Appl
30	271.5	9.5	524	9	US-09-864-761-38213	Sequence 38213, A
31	271.5	9.5	524	15	US-10-104-047-3371	Sequence 3371, A
32	259	9.1	379	16	US-10-615-659-6	Sequence 6, Appli
33	259	9.1	379	16	US-10-635-977-6	Sequence 6, Appli
34	258	9.0	92	15	US-10-424-599-262294	Sequence 262294,
35	254.5	8.9	377	15	US-10-210-130-118	Sequence 118, App
36	254.5	8.9	377	15	US-10-250-613-1	Sequence 1, Appli
37	237	8.3	42	16	US-10-615-659-24	Sequence 24, Appl
38	237	8.3	42	16	US-10-635-977-24	Sequence 24, Appl
39	207	7.3	698	15	US-10-108-260A-4312	Sequence 4312, Ap
40	207	7.3	767	18	US-10-450-763-41893	Sequence 41893, A
41	205.5	7.2	553	15	US-10-108-260A-2495	Sequence 2495, Ap
42	205.5	7.2	887	16	US-10-473-574-20	Sequence 20, Appl
43	186	6.5	663	16	US-10-322-281-133	Sequence 133, App
44	183	6.4	404	15	US-10-104-047-3621	Sequence 3621, Ap
45	178.5	6.3	237	18	US-10-450-763-41569	Sequence 41569, A

ALIGNMENTS

RESULT 1

US-10-615-659-2
; Sequence 2, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-2

Query Match	100.0%	Score	2854	DB	16	Length	541
Best Local Similarity	100.0%	Pred. No.	38-229				
Matches	541	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MASSILKVVVSHQSCSRSSR	QDQREAEAGSDLSRQDAENAEAKRLGPGQVLDIAC	60			
Db	1	MASSILKVVVSHQSCSRSSR	QDQREAEAGSDLSRQDAENAEAKRLGPGQVLDIAC	60			
Qy	61	KVQCAVYLGQLEHEDITSADAVE	DLTEAEWEDLTQQYSLVHGDAFTSNRNYFSQCOAL	120			
Db	61	KVQCAVYLGQLEHEDITSADAVE	DLTEAEWEDLTQQYSLVHGDAFTSNRNYFSQCOAL	120			
Qy	121	LNRTSNPOTDIDGLRNWIIPAKS	RGRDIVCMRVEEILELAAADHPLSRDNKVVV	180			

Db 121 LNRITSVNPTDIDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAAADHPLSRDNKVV 180
QY 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSQRFSLDKLDSAIHLCNN 240
Db 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSQRFSLDKLDSAIHLCNN 240
QY 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
Db 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDRFRPMLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDRFRPMLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
QY 361 RSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
QY 421 ARGPSAMPDPAQPPSPALQORDLGLKEEGLPLALLAPLRGAESGGAQPTTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQORDLGLKEEGLPLALLAPLRGAESGGAQPTTKAAGKV 480
QY 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLKLTAEAGLRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLKLTAEAGLRPPPGKG 540
QY 541 S 541
Db 541 S 541

RESULT 2
US-10-635-977-2
; Sequence 2, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; PRIOR FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S. 10/615,659
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-2

Query Match 100.0%; Score 2854; DB 16; Length 541;
Best Local Similarity 100.0%; Pred. No. 3e-229; Indels 0; Gaps 0;
Matches 541; Conservative 0; Mismatches 0

QY 1 MASSILKVVVSHQSCSRSSKPRDQREAGSSDLSSRQDAENAEAKRLGLPQLVDIAC 60
Db 1 MASSILKVVVSHQSCSRSSKPRDQREAGSSDLSSRQDAENAEAKRLGLPQLVDIAC 60
QY 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNSRNYFSQCQAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNSRNYFSQCQAL 120
QY 121 LNRITSVNPTDIDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAAADHPLSRDNKVV 180
Db 121 LNRITSVNPTDIDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAAADHPLSRDNKVV 180
QY 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSQRFSLDKLDSAIHLCNN 240
Db 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSQRFSLDKLDSAIHLCNN 240

Db 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSQRFSLDKLDSAIHLCNN 240
QY 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
Db 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDRFRPMLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFVLGRDRFRPMLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360
QY 361 RSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWRQPVVEPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
QY 421 ARGPSAMPDPAQPPSPALQORDLGLKEEGLPLALLAPLRGAESGGAQPTTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQORDLGLKEEGLPLALLAPLRGAESGGAQPTTKAAGKV 480
QY 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLKLTAEAGLRPPPGKG 540
Db 481 ELPACPCRHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLKLTAEAGLRPPPGKG 540
QY 541 S 541
Db 541 S 541

RESULT 3
US-10-615-659-13
; Sequence 13, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: TYROSINE-LIGASE-LIKE PROTEIN, BGS42
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-13

Query Match 99.6%; Score 2843; DB 16; Length 541;
Best Local Similarity 99.6%; Pred. No. 2.5e-228;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MASSILKVVVSHQSCSRSSKPRDQREAGSSDLSSRQDAENAEAKRLGLPQLVDIAC 60
Db 1 MASSILKVVVSHQSCSRSSKPRDQREAGSSDLSSRQDAENAEAKRLGLPQLVDIAC 60
QY 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNSRNYFSQCQAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAEWEDLTQOYYSLVHGDAFISNSRNYFSQCQAL 120
QY 121 LNRITSVNPTDIDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAAADHPLSRDNKVV 180
Db 121 LNRITSVNPTDIDGLRNIIWKPAKSRGRDIVCMDRVVEEILELAAADHPLSRDNKVV 180
QY 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSQRFSLDKLDSAIHLCNN 240
Db 181 QKIETPLLICDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSQRFSLDKLDSAIHLCNN 240
QY 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
Db 241 AVQYILKNDVGRSPLLPAAHNMWTSRFOEYLORQGRGAVMGSVIYPSMKKAIHAHMKVAQ 300
QY 301 DHVEPRKNSFELYGADFVLGRDRFRPMLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVD 360

Db 301 DHVEPRKNSFELYGADFLVGRDFPWLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVD 360
Qy 361 RSCDIGNFELLWRQPVVVEPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
Db 361 RSCDIGNFELLWRQPVVVEPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQPPSPALQDGLGKEKGILPALLAPLRGAABSGGAAQPTRTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQDGLGKEKGILPALLAPLRGAABSGGAAQPTRTKAAGKV 480
Qy 481 ELPACPCRHHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGKLTAEAGALRPPGKG 540
Db 481 ELPACPCRHHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGKLTAEAGALRPPGKG 540
Qy 541 S 541
Db 541 S 541

RESULT 4

US-10-635-977-13
; Sequence 13, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; CURRENT FILING DATE: 2003-08-07
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-13

Query Match 99.6%; Score 2843; DB 16; Length 541;

Best Local Similarity 99.6%; Pred. No. 2.5e-228;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MASSILKVVVSHQSCSRSSKPRDQREAGSSDLSSRQDAENAEAKLRLPGQLVDIAC 60
Db 1 MASSILKVVVSHQSCSRSSKPRDQREAGSSDLSSRQDAENAEAKLRLPGQLVDIAC 60
Qy 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCCAL 120
Db 61 KVCQAYLGQLEHEDIDTSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCCAL 120
Qy 121 LNRTSVNPQTDIGLNIWIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKVV 180
Db 121 LNRTSVNPQTDIGLNIWIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKVV 180
Qy 181 QKIYETPLLCIDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLSAHLN 240
Db 181 QKIYETPLLCIDTKFDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLSAHLN 240
Qy 241 AVQKYLKNDVGRSPLLPANHNTSTRFQYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQ 300
Db 241 AVQKYLKNDVGRSPLLPANHNTSTRFQYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQ 300
Qy 301 DHVEPRKNSFELYGADFLVGRDFPWLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVD 360
Db 301 DHVEPRKNSFELYGADFLVGRDFPWLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVD 360
Qy 361 RSCDIGNFELLWRQPVVVEPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420

Db 361 RSCDIGNFELLWRQPVVVEPPFSGSDLCVAGSVRRARRQVLPVCNLIKASASLLDAQPLK 420
Qy 421 ARGPSAMPDPAQPPSPALQDGLGKEKGILPALLAPLRGAABSGGAAQPTRTKAAGKV 480
Db 421 ARGPSAMPDPAQPPSPALQDGLGKEKGILPALLAPLRGAABSGGAAQPTRTKAAGKV 480
Qy 481 ELPACPCRHHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGKLTAEAGALRPPGKG 540
Db 481 ELPACPCRHHVDSQAPNTGVPVPAQPAKSWDPNQLNAHPLEPVLRLGKLTAEAGALRPPGKG 540
Qy 541 S 541
Db 541 S 541

RESULT 5

US-10-615-659-14
; Sequence 14, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-14

Query Match 54.7%; Score 1560; DB 16; Length 293;

Best Local Similarity 100.0%; Pred. No. 1.4e-121;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 EDIDTSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCCALLNRITSVNPQTD 132
Db 1 EDIDTSADAVEDLTEAWEDLTQYYSLVHGDAFISNSRNYFSQCCALLNRITSVNPQTD 60
Qy 133 IDGLRNIWIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKVVQKIYETPLICD 192
Db 61 IDGLRNIWIKPAKSRGRDIVCMRVVEIELEAAADHPLSRDNKVVQKIYETPLICD 120
Qy 193 TKEDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLSAHLNNAVQKYLKNDVGR 252
Db 121 TKEDIRQWFLVTDWNPNTIWFYKESYLRFSTQRFSLDKLSAHLNNAVQKYLKNDVGR 180
Qy 253 SPLLPANHNTSTRFQYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 312
Db 181 SPLLPANHNTSTRFQYLRQGRGAVGWSVIYPSMKKAIHAHMKVAQDHVEPRKNSFEL 240
Qy 313 YGADFVLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVD RSCDI 365
Db 241 YGADFVLGRDFRPWLIEINSSPTMHPSTPTVAQLCAQVQEDTIKVAVD RSCDI 293

RESULT 6

US-10-635-977-14
; Sequence 14, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725

;
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-14

Query Match 54.7%; Score 1560; DB 16; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.4e-121;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 EDIDTADAVEDTEAEWEDLTQOYSLVHGDAFISNRYFSQCCALLNRITSVNPQTD 132
DB 1 EDIDTADAVEDTEAEWEDLTQOYSLVHGDAFISNRYFSQCCALLNRITSVNPQTD 60

QY 133 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILEAAADHPLSRDNKVVQKYIETPLLCD 192
DB 61 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILEAAADHPLSRDNKVVQKYIETPLLCD 120

QY 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQSFSLDKLDSAIHLCNNAVQKYLKNDVGR 252
DB 121 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQSFSLDKLDSAIHLCNNAVQKYLKNDVGR 180

QY 253 SPLLPAHNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
DB 181 SPLLPAHNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 240

QY 313 YGADFVLGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDNRSCDI 365
DB 241 YGADFVLGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDNRSCDI 293

RESULT 7
US-10-615-659-4
; Sequence 4, Application US/10615659
; Publication No. US20040157234A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283 NP
; CURRENT APPLICATION NUMBER: US/10/615,659
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-615-659-4

Query Match 45.5%; Score 1299; DB 16; Length 242;
Best Local Similarity 100.0%; Pred. No. 6.8e-100;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILEAAADHPLSRDNKVVQKYIETPLLCD 192
DB 1 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILEAAADHPLSRDNKVVQKYIETPLLCD 60

QY 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQSFSLDKLDSAIHLCNNAVQKYLKNDVGR 252
DB 61 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQSFSLDKLDSAIHLCNNAVQKYLKNDVGR 120

QY 253 SPLLPAHNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
DB 121 SPLLPAHNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180

QY 313 YGADFVLGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDNRSCDI 372
DB 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDNRSCDI 240

QY 373 RQ 374
DB 241 RQ 242

RESULT 8
US-10-635-977-4
; Sequence 4, Application US/10635977
; Publication No. US20040171131A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL TESTIS-SPECIFIC TUBULIN
; FILE REFERENCE: D0283A CIP
; CURRENT APPLICATION NUMBER: US/10/635,977
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/394,725
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: U.S.10/615,659
; PRIOR FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-977-4

Query Match 45.5%; Score 1299; DB 16; Length 242;
Best Local Similarity 100.0%; Pred. No. 6.8e-100;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILEAAADHPLSRDNKVVQKYIETPLLCD 192
DB 1 IDGLRNIWIIPKAASGRGRDIVCMRDRVEEILEAAADHPLSRDNKVVQKYIETPLLCD 60

QY 193 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQSFSLDKLDSAIHLCNNAVQKYLKNDVGR 252
DB 61 TKFDIRQWFLVTDWNPPLTIWFYKESYLRSTQSFSLDKLDSAIHLCNNAVQKYLKNDVGR 120

QY 253 SPLLPAHNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 312
DB 121 SPLLPAHNMTSTRFOEYLQROGRGAVGWSVIYPSMKKAIAHAMKVAQDHVEPRKNSFEL 180

QY 313 YGADFVLGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDNRSCDI 372
DB 181 YGADFVLGRDPRFWLIEINSSPTMHPSTPTVTAQLCAQVQEDTIKVAVDNRSCDI 240

QY 373 RQ 374
DB 241 RQ 242

RESULT 9
US-10-275-595A-5
; Sequence 5, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil

This Page Blank (uspto)